

# 测量系统分析

参考手册

第四版

## 前言

本参考手册是在汽车工业行动集团的赞助下，由克莱斯勒集团公司、福特汽车公司和通用汽车公司供方质量要求特别工作组认可的测量系统分析（MSA）工作小组编写。负责第四版的工作小组成员是 Michael Down(通用汽车公司)、Frederick Czubak(克莱斯勒集团公司)、Gregory Gruska(奥曼克公司)、Steve Stahley(康明斯公司)与 David Benham.

本手册是对测量系统分析的一种介绍，它并不试图去限制某特定过程或商品所适用的分析方法的发展。当这些指导文件意在涵盖测量系统通常发生的情况时，其中可能还有一些问题没有考虑到；这些问题应该直接反馈给您的受权顾客代表。

本手册的版权归克莱斯勒集团公司、福特汽车公司和通用汽车公司。保留所有权利 2010.

2010 年 6 月

## MSA 第四版快速指南

测量系统类型	MSA 方法	章节
基本计量型	级差, 均值和极差, 方差分析 (ANOVA), 偏倚, 线性, 控制图	三
基本计数型	信号探测法, 假设试验分析法	三
不可重复 (例如, 破坏试验)	控制图法	四
复杂计量型	极差法, 均值和极差法, 方差分析 (ANOVA) 法, 偏倚, 线性, 控制图法	三、四
复合的系统, 量具 或试验标准	控制图法, 方差分析 (ANOVA), 回归分析法	三、四
其他	替代的方法	五
其它	白皮书---可上网查询, 网址 <a href="http://www.aiag.org">http://www.aiag.org</a>	

注: 关于 GRR 标准差的使用

## 目 录

<b>第一章</b>	<b>测量系统总指南</b>
<b>第 A 节</b>	引言、目的及术语
	引言
	目的
	术语
<b>第 B 节</b>	测量过程
	测量系统
	测量系统变差的影响
<b>第 C 节</b>	测量策划和计划
<b>第 D 节</b>	测量资源的开发
	量具资源选择过程
<b>第 E 节</b>	测量问题
<b>第 F 节</b>	测量不确定度
<b>第 G 节</b>	测量问题分析
<b>第二章</b>	<b>用于评估测量系统的基本概念</b>
<b>第 A 节</b>	背景
<b>第 B 节</b>	选择/开发试验程序
<b>第 C 节</b>	测量系统研究的准备
<b>第 D 节</b>	结果分析
<b>第三章</b>	<b>对可重复测量系统推荐的实施方法</b>
<b>第 A 节</b>	试验程序范例
<b>第 B 节</b>	计量型测量系统研究- 指南
	用于确定稳定性的指南
	确定偏倚的指南- 独立样本法
	确定偏倚的指南- 控制图法
	确定性的指南
	确定重复性和再现性的指南
	极差法
	平均值和极差法
	方差分析法 (ANOVA)
<b>第 C 节</b>	计数型测量系统研究
	风险分析法
	信号检查 (signal detection) 方法
	分析方法
<b>第四章</b>	<b>其他测量概念和实践</b>
<b>第 A 节</b>	不可重复的测量系统的实践
<b>第 B 节</b>	稳定性研究
<b>第 C 节</b>	变差研究
<b>第 D 节</b>	识别过大的零件内部变差的影响

<b>第 E 节</b>	平均值和极差法—额外的处理
<b>第 F 节</b>	量具性能曲线
<b>第 G 节</b>	通过多次读值减少变差
<b>第 H 节</b>	聚焦标准差法计算 GRR
<b>附录</b>	
<b>附录 A</b>	方差分析的概念
<b>附录 B</b>	GRR 对能力指数 $C_p$ 的影响 公式 分析 图形分析
<b>附录 C</b>	
<b>附录 D</b>	量具 R 研究
<b>附录 E</b>	用误差修正术语替代 PV 计算
<b>附录 F</b>	P. I. S. M. O. E. A 误差模型
<b>术语</b>	
<b>参考文献</b>	
<b>范例表格</b>	
<b>索引</b>	

# 第一章

## 测量系统总指南

### 第一章---第 A 节

#### 引言、目的及术语

#### 引言

测量数据的使用比以前更多更广泛了。例如，现在是否对制造过程进行调整的决定通常以测量数据为基础，将测量数据或一些从它们所计算出的统计值，与这一过程的统计控制限（statistical control limits）相比较，如果该比较过程已超出统计控制，则进行某种调整，否则，该过程将被允许在没有调整的状态下运行。测量数据的另一个用处是确定在两个或更多变量之间是否存在显著的相互关系。例如，如果怀疑一个模塑零件上的一个关键尺寸和注射材料的温度有关。这种可能的关系可以通过采用所谓回归分析的统计方法来研究，即比较关键尺寸的测量值和注射材料的温度测量值

进行这种举例我互相关系探测研究，被戴明博士称为分析研究法。通常，分析研究法是不断增加对影响过程系统原因知识的一种分析研究。分析研究是测量数据和最重要应用之一，因为应用它们能使得对过程有更好的理解。

使用以数据为基础的程序的最大益处取决于所使用的测量数据的质量。如果测量数据质量低，程序的益处可能会较低。同样的，测量数据质量高，收益也可能较高。

为了确保应用测量数据所得到的收益大到足以承担获得这些数据的成本，数据的质量需要特别的注意。

#### 测量数据的质量

数据的质量取决于从处于稳定条件过户进行操作的测量系统中，多次测量的统计特性。例如，假定使用某一在稳定条件下操作的测量系统对某一特定特性进行了几次测量。如果这些测量值均与该特性的参考值“接近”，那么可以说这些测量数据的质量“高”，同样，如果部分或所有的测量值与参考值相差“很远”，则称数据的质量“低”。

表征数据质量最通用的统计特性是测量系统的偏倚和方差。所谓偏倚的特性，是指数据相对参考（基准）值的位置，而被称为变差的特性是指数据的分布宽度。

低质量数据最普通的原因之一是变差太大。一组数据中的变差多是由于测量系统及其环境相

互作用造成的。例如，一个用来测量一罐液体容积的测量系统，可能对该测量系统所处的环境中的大气温度较敏感。在这种情况下，数据的变差可能是因为环境温度变化造成的。因此，对测量的数据很难解释，因此，该测量系统不尽理想。

如果交互作用产生变差过大，那么数据的质量会很低，从而造成测量数据无法利用。例如，一个具有大量变差的测量系统，在分析制造过程中使用是不适合的，因为测量系统变差可能会掩盖制造过程的变差。管理一个测量系统的许多工作是监视和控制变差。其它的还需要把重点集中在了解测量系统与其环境有什么样的相互作用，以便获得可接受质量的数据。

## 目的

本手册的目的是为评定测量系统的质量提供指南。尽管这些指南足以用于任何测量系统，但主要用于工业界的测量系统。本手册不打算作为所有测量系统的一种分析总览，而是主要用于那些注每个零件能重复读数的测量系统。许多分析对于其它形式的测量系统也是很有用的，并且该手册的确包含了参考意见和建议，但对更复杂的或不常用的方法在此没有讨论，建议使用者参考适宜的统计资源。本手册也不涵盖顾客对测量系统分析方法所要求的批准。

## 术语

如果不建立一套术语来引述共同的统计特征和相关的测量系统要项，那么讨论测量系统的分析可能会造成混淆和误解。本节将用于本手册的术语汇总如下。

在本手册中使用了以下术语：

- 测量被定义为“对某具体事物赋予数字（或数值），以表示它们对于特定特性之间的关系。这个定义由C. Eisenhart（1963）首次提出。赋予数字的过程被定义为测量过程，而数值的指定被定义为测量值。
- 量具：是指任何用来获得测量的装置，特别是经常用在工厂现场的装置，包括通过/止规。
- 测量系统：是对测量单元进行量化或对被测的特性进行评估，其所使用的仪器或量具、标准、操作、方法、夹具、软件、人员、环境及假设的集合；也就是说，用来获得测量结果的整个过程。

由以上这些定义，可以将测量过程看成是一个制造过程，其产生的输出就是数值（数据）。这样看待一个测量系统是有用的，因为这样让我们明白已经说明的所有概念、原理和工具，这在统计过程控制中早已被证实他们的作用。

## 术语汇总<sup>1</sup>

### 标准（standard）

- 用于比较的可接受偏倚
- 验收标准
- 一已知的值，在不确定度（uncertainty）的指南范围内，被接受而为一真值（true value）
- 参考值（referencr value）

标准应该是一个可操作的定义：该定义在由供方或顾客应用时，将会产生同样的结果，并且在过去、今天、将来都有同样的含义。

## 基本的设备(basic equipment)

- **分辨力(discrimination)、可读性(readability)、解析度(resolution)**

- √ 别名：最小可读单位、测量解析度、最小刻度极限、或探测的最小极限
- √ 由设计所确定的固有特性
- √ 一个测量仪器或输出的最小刻度单位
- √ 通常被显示为测量单位
- √ 10: 1 的比例法则

- **有效解析度(effective resolution)**

- √ 特定应用条件下，一个测量系统对过程变差的灵敏度
- √ 可以导致测量有用的输出信号的最小输入
- √ 通常被描述为一种测量单元

- **参考值(reference value)**

- √ 某一个物品的可接受数的值
- √ 需要一个可操作的定义
- √ 常被用来替代真值使用

- **真值(true value)**

- √ 某一物品的真实数值
- √ 不可知且无法知道的

---

<sup>1</sup>见第一章第五节术语定义和讨论

## 位置变差 (location variation)

- **准确度(accuracy)**
  - √ 与真值或可接受的参考值“接近”的程度
  - √ 在 ASTM 包括了位置及宽度误差的影响
- **偏倚 (bias)**
  - √ 观测到测量的平均值与参考值之间的差值
  - √ 是测量系统的系统误差所构成
- **稳定性 (stability)**
  - √ 随时间变化的偏倚值
  - √ 一个稳定的测量过程在位置方面是处于统计上受控状态
  - √ 别名：漂移 (drift)
- **线性 (linearity)**
  - √ 在量具正常工作量程内的偏倚变化量
  - √ 多个独立的偏倚误差在量具工作量程内的关系
  - √ 是测量系统的系统误差所构成

## 宽度变差(width variation)

- **精密度<sup>2</sup>(precision)**
  - √ 每个重复读数之间的“接近”程度
  - √ 是测量系统的随机误差所构成
- **重复性**
  - √ 一个评价人多次使用一件测量仪器，对同一零件的某一特性进行多次测量下的变差
  - √ 是在固定和已定义的测量条件下，连续（短期）多次测量中的变差
  - √ 通常指 E. V. -设备变差(equipment variation)
  - √ 设备（量具）能力或潜能
  - √ 系统内部变差

---

<sup>2</sup>在 ASTM 文件中，没有测量系统的精密度这样的说法；也就是说，精密度不能用单一数值表述。



- **再现性**
  - √ 由不同的评价人使用相同的量具，测量一个零件的一个特性的测量平均值的变差。
  - √ 在对产品和过程进行鉴定时，误差可能是评价人、环境（时间）或方法
  - √ 通常指 A. V- 评价人变差(appraiser variation)
  - √ 系统间（条件）误差
  - √ 在 ASTM E456-96 包括：重复性、实验室、环境及评价人影响
  
- **GRR 或量具的重复性和再现性（ gage R & R）**
  - √ 量具的重复性和再现性；测量系统重复性和再现性联合估计值
  - √ 测量系统能力；取决于所用的方法，可能包括或不包括时间影响
  
- **测量系统性能(measurement system performance)**
  - √ 测量系统变差的短期估计值（例：“GRR”，也包括图表法）
  
- **灵敏度（sensitivity）**
  - √ 能导致可探测到的输出信号的最小输入
  - √ 测量系统对被测特性变化的感应度
  - √ 取决于量具设计（分辨力）、固有质量（OEM）使用期间的维修，以及测量仪器与标准的操作情况
  - √ 通常被描述为测量单元
  
- **一致性（consistency）**
  - √ 随时间重复性变化的程度
  - √ 一致的测量过程是在宽度（变差）方面处于统计上受控状态
  
- **均一性**
  - √ 整个正常操作范围内重复性的变化
  - √ 重复性的同义词



## 系统变差 (system variation)

测量系统变差可以分为:

- **能力 (capability)**
  - √ 短期获取读数的变异性
- **性能(performance)**
  - √ 长期读数的变化量
  - √ 以总变差 (total variation) 为基础
- **不确定度(uncertainty)**
  - √ 有关被测值的数值估计范围, 相信真值包括在此范围内

测量系统必须稳定和一致

测量系统总变差的所有特性均假设系统是稳定和一致的。例如, 变差分量可以包括第 14 页图 2 报示的各项的合成。

标准和可追溯性

国家标准和技术协会 (National Institute of Standards and technology, NIST) 是美国的主要的国家测量协会

(National measurements institute, NMI), 它属于美国商业部 (U.S. department of commerce),

NIST 的前身是国家标准局 (NBS), 是美国度量衡的最高权力机构。NIST 的主要责任是提供测量服务以及保存测量标准, 帮助美国工业界建立可溯源的测量, 最终能为产品和服务的贸易提供帮助。NIST 直接对许多类型的工业提供服务, 但主要是那些需要高准确度产品以及在过程中结合目前测量科技进步水准的行业提供服务。

国家测量协会  
(NMI)

世界上许多工业化国家维持了他们与 NIST 相似的 NMI, 也为他们的相关行业提供高水平的度量衡或测量服务。NIST 与其他国家的 NMI 合作研究, 从而确保在一个国家进行的测量不会与另一个国家不同。该项工作通过各个 NMI 之间的互相认可协定 (mutual recognition arrangements, MRAs), 并进行互相实验比较来完成。有一点要注意, 这些各国的 NMI 的能力有差别, 并且不是所有类型的测量均在一个共同目标的基础上进行比较, 所以存在着差别。这就是为什么要了解追溯到是谁的测量, 以及是如何进行追溯的重要性。

可追溯性 (Traceability)

对商品和服务的贸易中溯源性是一个非常重要概念。测量可以追溯到相同或类似的标准, 比不能追溯的测量更容易达成互相承认。可追溯的测量还可帮助减少重新试验的要求, 以及好产品的拒收与坏产品的接收。

可追溯性在 ISO 国际基本和通用的度量衡术语词汇 (ISO International Vocabulary of Basic and General Terms in Metrology, VIM) 中的定义为:

“通过一个完整的比较链追溯到规定的参考标准 (通常为家或国际标准) 的测量特性或标准值, 都具有一定的不确定度。”

建立一个测量的可追溯性一般可以通过一条比较链追溯到 NMI。但是在工业界的许多情况下, 测量的可追溯性可能追溯到顾客和供方同意的参考值或“一致的标准”。这些达成一致的标准和 NMI 之间的可追溯性也许不是总能清晰地被理解的, 所以, 最关键的是测量能追溯到满足顾客需求的程度。随着测量系统在工业界中的测量技术提升和目前科技进步水准, 测量追溯到哪里和如何追溯的定义将成为一个不断进化的概念。

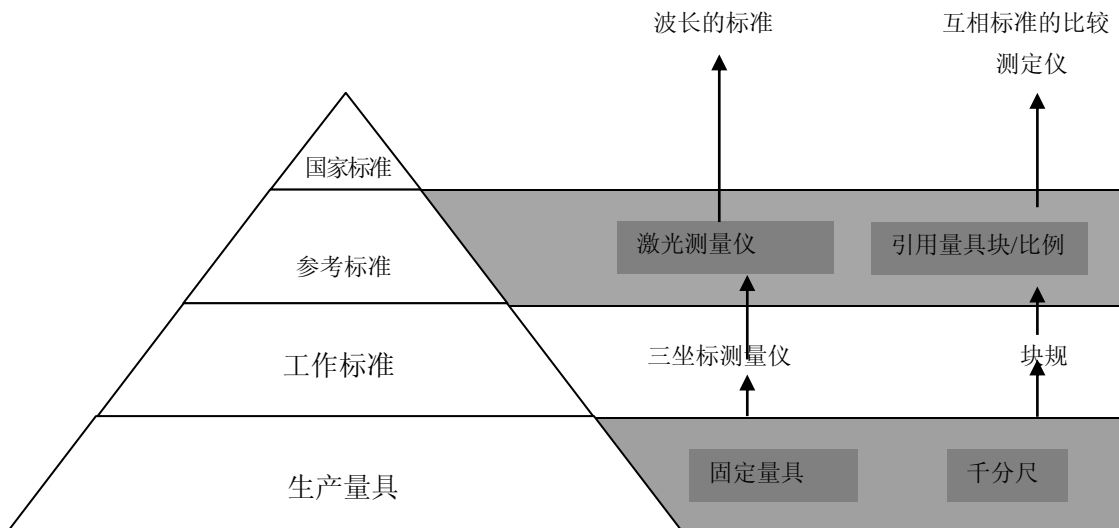


图 1: 长度测量溯源性链的示例

NMI 与不同的国家实验室、量具供方、科技开发制造商等密切合作, 从而确保他们的参考标准, 这些政府和私人工业机构然后可以使用他们的标准, 来为他们的客户的度量衡或量具实验室、校准工作或其他的首级标准提供校准和测量服务。这些情况的连接或链将应用在工厂的现场, 并提供基本的测量可追溯性。这个可以通过这种完整的测量链回溯到 NIST 的测量被称为可追溯到 NIST。

不是所有组织都有自己的内部计量或量具实验室, 因此, 他们必须依靠外部的商业/独立实验室提供可追溯性的校准或测量服务。如果商业/独立实验室的能力已通过实验室认可来保证, 以上方法是获得可追溯到 NIST 的可接受和适应的方法。

---

## 校准系统

校准系统指在特定环境下以建立测量设备与已知的参考价值和不确定值的可追溯标准之间关系的一套操作系统。校准系统同时也包括通过对与测量设备精度的误差调整来检测校准系统通过利用校准方法及标准来确定测量系统的测量的可追溯。

可追溯性是一个可以溯源到适当计量能力和测量不确定度的校准标准的一个校准链接事件。每个校准项目都由必备部分构成，包括校准标准、验证的测量和测试设备、校准方法及步骤、校准记录及合格工作人员。

单位应有拥有内部校准实验室或拥有能控制及维护校准项目的所有组成部分的部门。其这些内部实验室应拥有实验室范围内所列之具体的能进行校准的设施及设备能用于操作校准的方法及程序。

校准系统是部门质量管理体系的一个组成部分。因此，所有内部审计要求应包括校准系统在内。

测量保证方案（MAPs）能检验在整个校准系统中的测量流程的可接受性。测量保证方案一般包括通过对相同特性或参数的二次测量过程中的可追溯性是来源于校准过程中在进行初始量测的一个单独的校准环节。测量保证方案同时还包括利用统计程序控制系统（SPC）追踪测量过程中的长期稳定性。

注：ANSI/NCSL Z540.3 及 ISO10012 均能提供校准系统许多部件的型号。

外部、商业性质的独立校准服务供应商在进行校准项目时，其校准系统必须通过 ISO/IEC 17025 认证的验证。在无合格的实验室情况下，对于有些设备校准服务，由设备制造者承担。

---

## 真值

测量过程的目标是零件的“真”值，希望任何个别的读值能尽可能的（经济的）与真值接近。遗憾的是真值永远不能确切地得到。然而基于作业上完善地定义一个特征的参考值，以及使用具有产生更高解析度并可追溯到 NIST 的测量系统，能够使这种不确定度减到最低。因为参考值常被用于对真值的替代，这些术语通常被互换使用；但不建议这种互换使用。

## 第一章----第 B 节 测量过程<sup>3</sup>

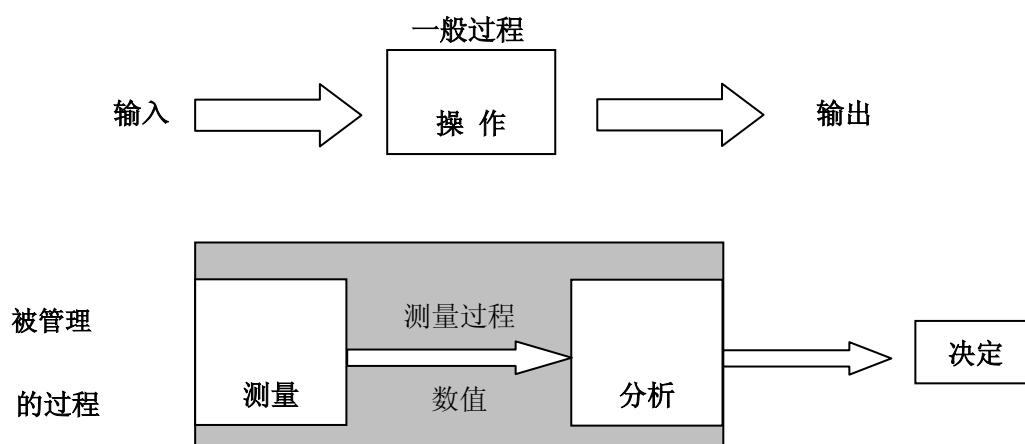
为了有效地管理任何过程的变差，需要以下知识：

- 过程应该做什么
- 会出什么错
- 过程正在做什么

规范和工程要求决定过程应该做什么。

过程失效模式及后果分析<sup>4</sup>（PFMEA）是用来确定与潜在过程失效相关的风险，并在这些失效出现前提出纠正措施。PFMEA 的结果转移至控制计划。

通过评价过程结果或参数，可以获得过程正在做什么的知识。这种活动，通常称为检验，是用适当的标准和测量装置，检查过程参数，过程中零件，已装配的子系统，或者是已完成的成品活动。这种活动能使观测者确定或否认过程是以稳定的方式操作并具有对顾客规定的目标而言可接受的变差这一前提。这种检查行为本身就是过程。



不幸的是，工业界传统上视测量和分析活动为“黑箱作业”。设备是主要关注点 – 特性越“重要”，量具越昂贵。很少顾虑仪器的适用性。因此这些量具通常没有得到适当的使用或简单地不去用它。

测量和分析活动是一个过程 – 测量过程。所有的过程控制管理，统计或逻辑技术均能应用。

这就意味着必须首先确定顾客和他们的需要。顾客，过程的拥有者，希望用最小的努力做出正确的决定。管理者必须提供资源以采购必要的和足够的设备来完成工作，但是采购最好的或最新的测量技术未必能保证做出正确的生产过程控制决定。

<sup>3</sup> 本章的部分内容经允许采用了《测量系统分析》指南，由 G.F.ruska 和 M.S.Heaphy 编写，第三代，1987,1988。

<sup>4</sup> 参见潜在的失效模式及后果分析（FMEA）参考手册-第 3 版

设备只是测量过程的一部分,过程的拥有者必须了解如何正确使用这些设备及如何分析和解释结果。因此管理者也必须提供清晰明了的操作规定和标准以及培训和支持。接下来,过程的拥有者有义务监控和控制测量过程,以确保获得稳定和正确的结果,包括整个测量系统分析方法- 对量具、程序、使用者及环境的研究,例如,正常操作条件的研究。

### 测量系统的统计特性

一个理想的测量系统在每次使用时均能产生“正确”的测量结果。每个测量都会遵循某个标准<sup>5</sup>。能产生这样的测量结果的测量系统被称为具有如下的统计特性:零方差、零偏倚和对所测的任何产品被错误分析的可能性为零。遗憾的是,具有这样理想统计特性的测量系统很少存在,因此过程管理者被迫使用统计特性不太理想的测量系统。测量系统的质量通常仅仅取决于经过一段时间后产生数据的统计特性。当他们要促成一整体上令人满意的测量系统时,其它因素如成本,易于操作等也同样重要。但是这是所产生数据的统计特性,是测量系统质量的决定因素。

对某种使用情况下非常重要的统计特性,其不一定是另一种使用情况下非常重要的统计特性。例如:在某些情况下使用三坐标测量仪(CMM),最重要的统计特性是“小”的偏倚和方差。具有这样统计特性的CMM将产生接近可追溯标准的公认值的测量结果。用这样仪器上所得到的数据对分析一个制造过程可能是十分有用的。但是,不管其偏倚和方差多么“小”,使用CMM这样的测量系统可能不能够用来分辨好产品或坏产品的接受工作,由于测量系统中其他因素会产生额外的变差来源。

---

<sup>5</sup>有关标准问题的完整讨论见《走出危机》,W. Edwards Deming,1982,1986,P. 279-281.

管理者有责任为最佳的数据应用，识别最为重要的统计特征；管理者也有责任确保使用这些特征作为选择测量系统的依据。为了达成以上目的，需要有关统计特性的具体定义，并且规定测量它们的可接受的方法。尽管每一个测量系统可能被要求有不同的统计特性，但有一些基本特性用于定义“好的”测量系统。包括：

- 1) 足够的分辨率和灵敏度。测量的增量应该小于测量的目的相应的过程变差或规范限值。通常被称为 10 比 1 规则，也就是说仪器的分辨力应把公差（或过程变差）划分成 10 等份或更多。这比例规则的意图是作为选择量具时的一个实际最先遵守的原则。
- 2) 测量系统应该是统计受控制状态<sup>6</sup>。这意味着在重复测量条件下，测量系统中的变差只能是由普通原因造成，而不能由特殊原因造成。这种情况可称之为具有统计的稳定性且可以通过控制图法最佳地进行评价。
- 3) 为了产品控制，测量系统的变异性与公差相比必须小。依据特性的公差评价测量系统。对于过程控制，测量系统的变差必须小于规范限值。以特性的公差来评估测量系统。
- 4) 为了过程控制，测量系统的变差应该能证明具有有效的解析度，并且小于制造过程的变差。 $6\sigma$  过程变差和/或 MSA 研究中的总变差可用来评估测量系统。

*测量系统的统计特性随着被测量的项目不同可能会发生变化。如果这样，测量系统的最大（最坏）的变差必须小于过程变差或规范限值。*

---

### 变差的来源

与所有过程类似，测量系统受到变差的随机原因和系统上变差的影响。这些变差来源由普通原因和特殊原因造成的。为了控制测量系统变差：

- 1) 识别潜在的变差源
- 2) 消除（如有可能的话）或监控这些变差的来源

尽管特殊原因取决于不同情况，但一些典型的变差源是可以被识别。有多种不同的方法来呈现与分类这些变差的来源，例如因果图、失效树等，但本指南将关注的是测量系统的主要要素。

<b>S</b>	标准
<b>W</b>	工作件(即零件)
<b>I</b>	仪器
<b>P</b>	人/程序
<b>E</b>	环境

缩写 S.W.I.P.E<sup>7</sup>用来表示一个普遍化的测量系统为了达成被要求的目标，有六个必要因素。S.W.I.P.E 分别代表标准、工作件、仪器、人和程序及环境。可以考虑作为整个测量系统的一个误差模型<sup>8</sup>。

---

<sup>7</sup> 这些缩写由度量衡专家 Ms.Mary 与 Honeywell,Eli Whitney Metrology Laband the Bendixcorporationg 首创提出

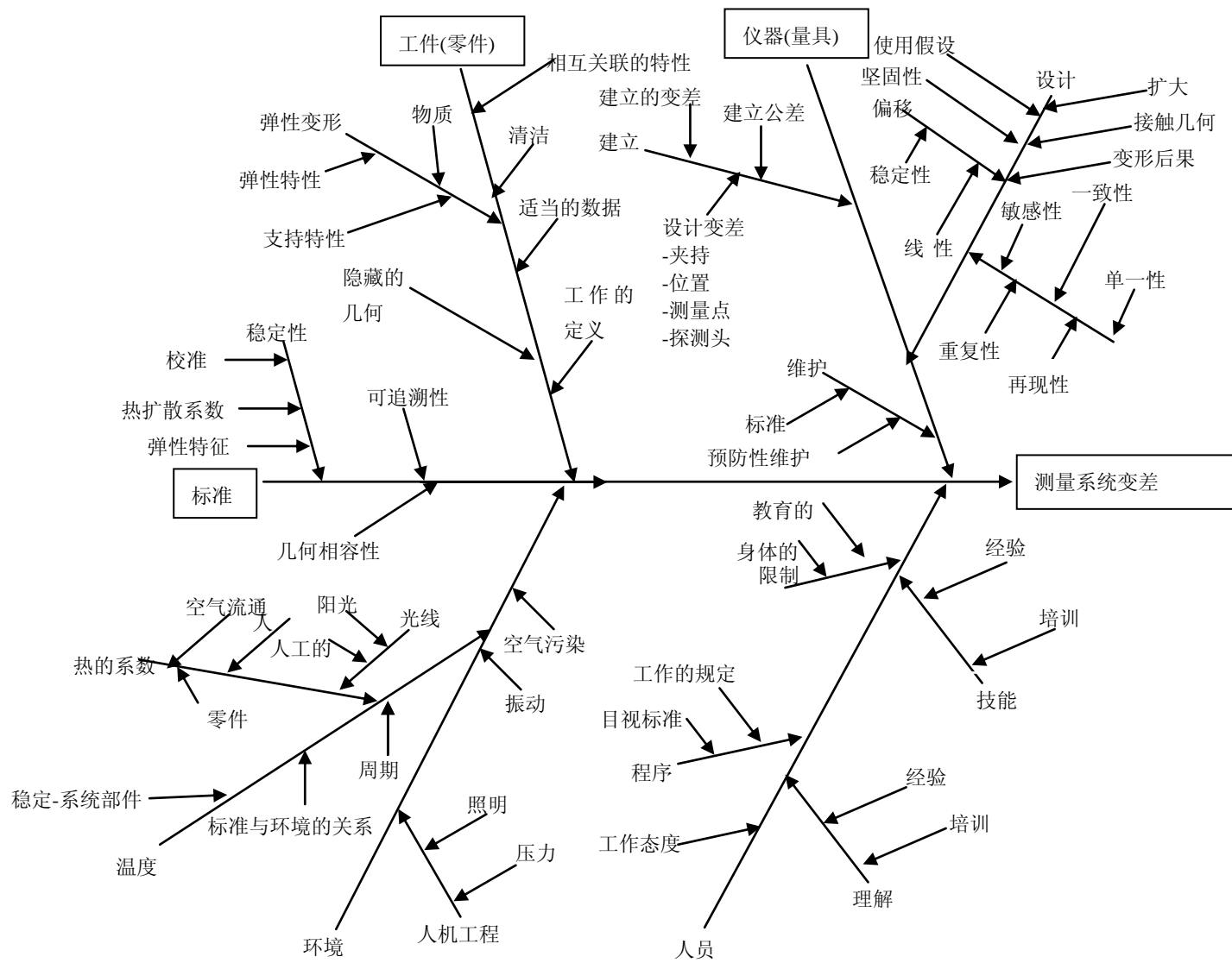
<sup>8</sup> 另一个供选择的误差模型见附件 F,P.I.S.M.O.E.A

需要理解影响那六个区域的因素，以便能够控制或消除它们。

图 2 显示了一张潜在的变差源的因果图。由于影响一个特定的测量系统的实际变差来源对该系统将是唯一的，所以，这图可作为开发一个测量系统变差来源的一个思考启动器。



图 2: 测量系统变差的因果图



## 测量系统变差的影响

由于测量系统会被不同的变差来源影响，相同零件的重复读值不会有相同的产出与结果，每个读值之间不同是因为普通和特殊原因变差造成的。

重复读数也不产生相同或同样的结果。读数之间不相同是由于普通和特殊原因造成。

对测量系统变差的不同变差来源影响应该进行经过短期和长期评价。测量系统的能力是测量系统在短期间的（随机）误差。它是由线性、均一性与再现性结合的误差量。测量系统性能，像过程性能一样，是所有变差来源在长期的影响。这是经由确定了我们的过程是否处于统计受控状态（如，稳定并一致；仅有普通原因的变差），对中性（无偏倚），以及具有期望结果在一定范围内的可接受变差，（量具重复性和再现性（GRR））的实现，这是在测量系统能力上增加了稳定性和一致性。

由于我是用测量系统的输出来对某产品和过程做出决策，这所有变差来源的累积后果通常被称为测量系统误差，或有时简单地称为“误差”。

## 对决策的影响

测量一个零件后，可以对零件状态的确定采取一个行动措施。一般来说，可以用来确定该零件是否可接受（在规范内）或不可接受（超出规范）。另一个常用的情况是将零件分类为特定的类别（例如：活塞的尺寸）

以一个范例作为接下来的讨论，有两种分类将被使用：超出规范“坏”和在规范内“好”；这并不为其它类别活动的讨论做出限制。

进一步的分类可能是可返工的、可补救的或报废。按照产品控制理论（product control philosophy），这样分类活动是测量一个零件的主要原因。但是，在过程控制理论下，集中的关注是零件的变差是否由过程中的普通原因还是特殊原因造成的。

**表 1：控制理论和驱使的关注点**

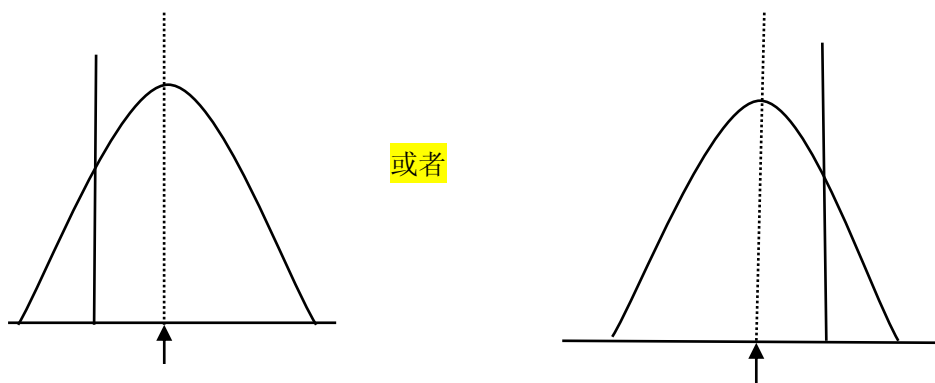
原理	关注点
产品控制	零件是否在指定范围内？
过程控制	过程变差是否稳定并可接受？

下一部分将针对测量误差对产品决策的影响。接下来再讨论它对过程决策的影响

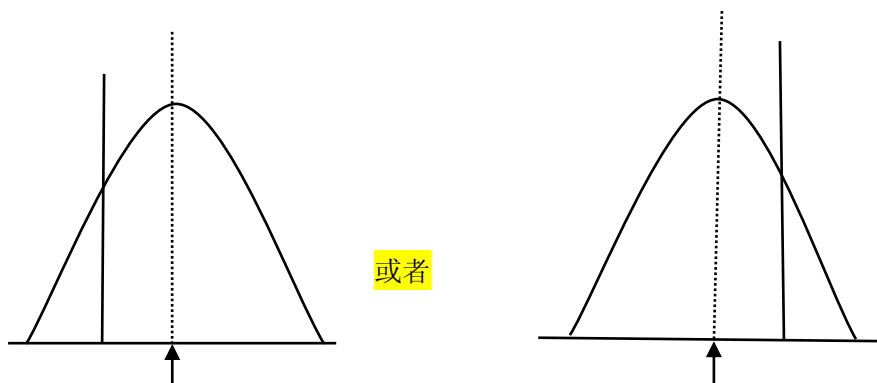
## 对产品决策的影响

为了更好地理解测量系统误差对产品决策的影响，考虑的案例是假设对某一个单一零件的多次读值的所有变差是由于量具的重复性和再现性引起的。也就是说，测量过程处于统计受控状态，要考虑单个零件重复读数所有变差由量具的重复性和再现性影响。那就是测量过程是统计受控状态，并且是零偏倚。

无论何时任何零件的测量分布区域与规范限值重迭时，有时可能会做出错误的决定。例如，如果出现以下情况，一个好的零件有时会被判为“坏”的（第 I 类误差，制造者风险或错误警告）

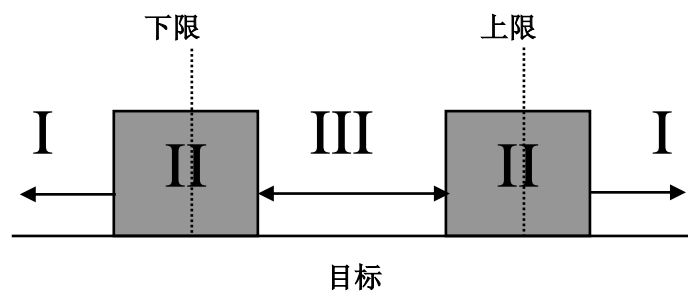


然后，如果出现以下情况，一个坏的零件有时会被判为“好”的（第 II 类误差，顾客的风险或过失比率），如果：



备注：错误警告比率+过失比率=误差比率

存在风险是作决策的机会，也可能会作出对个人或进程不利的决策。也就是说，考虑到规范限值，只要当测量系统误差与规范限值相交时，就可能发生对零件做出潜在的错误决定。以下提出三个明显的区域。



图中，

- 第 I 类区 坏零件永远被称为坏零件
- 第 II 类区 可能做出潜在的错误决定
- 第 III 类区 好零件永远被称为好零件

对于产品状况，目标是最大限度地做出正确决定，有以下两种选择：

- 1) 改进生产过程：减少过程的变差，没有零件产生在 II 区域。
- 2) 改进测量系统：减少测量系统误差从而减小 II 区域的面积，因此生产的所有零件将在 III 区域，这样就能减少做出错误决定的风险。

上述讨论假定测量过程是统计受控并且是对准目标。如果以上两个假设中的任何一个不成立，那么很难相信任何观测值能带来正确的决定。

### 对过程决策的影响

对于过程控制，以下需求要被建立：

- 统计的控制
- 对准目标
- 具有可接受的变差

正如上一部分的解释，测量误差可能造成对产品的错误决定。测量误差对过程决策的影响如下：

- 把普通原因识别为特殊原因
- 把特殊原因识别为普通原因

测量系统变异性可能影响过程的稳定性、目标以及变差的决定。实际和观测的过程变差之间的基本相互关系是：

$$\sigma^2_{\text{观}} = \sigma^2_{\text{实}} + \sigma^2_{\text{测}}$$

式中，

$$\begin{aligned}\sigma^2_{\text{观}} &= \text{观测到的过程变差值} \\ \sigma^2_{\text{实}} &= \text{实际过程变差} \\ \sigma^2_{\text{测}} &= \text{测量系统的变差}\end{aligned}$$

能力指数<sup>9</sup>C<sub>p</sub> 定义为

$$C_p = \frac{\text{公差范围}}{6\sigma}$$

---

<sup>9</sup> 虽然此处讨论使用 C<sub>p</sub>，也可能用性能指数 P<sub>p</sub> 来计算。

将上面的公式进行替代可以得出观测过程指数与实际过程指数之间的关系：

$$(Cp)^2_{\text{观}} = (Cp)^2_{\text{实}} + (Cp)^2_{\text{测}}$$

假定该测量系统处于统计受控状态而且对正目标，实际过程 Cp 可以与观测 Cp 的图形上比较大致相等<sup>10</sup>。

因此，观测的过程能力结合了实际过程能力加上测量过程造成的误差。要想达到某一特定的过程能力目标，需要考虑测量变差中的因素。

举例说明，如果测量系统 Cp 值是 2，实际过程 Cp 必须大于或等于 1.79，才能得到 1.33 的计算值（观测值）。如果测量系统 Cp 值本身只有 1.33，则该过程需要没有任何变差才能使观测到的过程能力达到 1.33—显而易见，这是一个不可能存在的情况。

## 新过程的接受

当一个新的过程中购买了如机械、制造、冲压、材料处理、热处理或装配，通常会有作为采购验收活动的一系列步骤需要完成；一般会在供方的场所进行设备研究，然后在顾客的场所进行研究。

在正常条件下使用时，如果测量系统在某个地方使用与在另一地方使用不一致时，可能会产生混淆。使用不同的测量仪器最常见情况是，在供方使用的仪器分辨率高过生产场地使用的仪器（量具）。例如，在采购验收期间使用坐标仪来测量零件，然后在生产期间用高度规测量；再如，在采购验收用一电子天平或实验室里的机械式天平作测量（称重），但在生产期间用简单的机械式天平作测量。

在这案例，当（较高分辨力）测量系统被使用在验收期间，其 GRR 为 10%，且实际过程 Cp 为 2.0，那么在验收期间观测到的过程 Cp 值将为 1.96<sup>11</sup>。

当这过程是在生产场地使用生产用的量具来研究时，会观测到更大的变差（也就是说一个较小的 Cp 值）。例如，如果生产量具的 GRR 为 30%，而实际过程能力 Cp 值还是 2.0，则观测到的过程能力 Cp 值将为 1.71。

最坏的情况是使用了不合格的生产用量具。如果测量系统的 GRR 是 60%（但是并不知道这一事实），那么观测到的 Cp 值将是 1.28。观测到的 Cp 值由 1.96 到 1.28 之间的区别就是由测量系统的不同所造成的。如果不能了解这情况，可能要努力地寻找新的过程在什么地方出错了，当然这些努力是白费的。

---

<sup>10</sup> 附录 B- 公式和图表。

<sup>11</sup> 对这个问题的讨论，假设没有样件的变差。事实上 1.96 是一个期望值，实际的结果会在其周围变化。

\*译者注：该公式为原文的，估计有误。

## 过程作业准备/控制（漏斗试验）

通常在制造场所会用每天开始工作时的第一个零件来验证过程是否对准目标值。如果被测零件没有对准目标值，则对过程进行调整；接下来，在某情形下另一个零件被测量，并且过程再度被调整。戴明博士将这种测量和做决定的方式称为“擅改（tampering）”

让我们思考一种情况：在某零件表面镀上一种贵金属的重量被控制在目标值 5.00 克。假设不知道用来测量重量的天平偏离了  $\pm 0.20$  克，因为从来没有对测量系统进行分析，所以人们不知道这一事实。作业指导书要求操作者在作业准备时每间隔 1 小时用一件产品进行重量验证。如果结果超过了 4.90 至 5.10 的范围，则操作者重新进行过程的作业准备。

在进行作业准备时，该过程在 4.95 克下运行，但由于测量误差操作者观测为 4.85 克。按照作业指导书的规定操作者设法将过程上调 0.15 克，目前这过程在目标值为 5.10 克的情况下生产。当操作者在下一个时间进行作业准备验证时，他观测到 5.08 克，于是过程被允许生产，可是这过程已经是被加上变差的过度调整（over-adjustment），并将会继续下去。

这是漏斗试验（funnel experiment）的一个例子，戴明博士称为为擅改<sup>12</sup>影响的结果。测量误差加重了这种问题。

漏斗试验有以下四个规则：

规则 1：除非过程不稳定，否则不要进行调整或采取措施。

规则 2：从过程上一次被测量偏向的相反方向，对过程调整一个相等的量。

规则 3：将过程重新设置到目标值，然后从目标值往相反方向调整一个相等的量。

规则 4：将过程调整到上一次测量的位置。

为贵金属镀层过程的作业准备指导书是规则 3 的一个范例。规则 2、3 和 4 将累进的增加更多变差。规则 1 是产生最小变差的最佳选择

漏斗试验的其它范例如下：

- 武断的以一个界限为基础来校准量具——也就是说界限值不能反映测量系统的变差（规则 3）
- 武断的使用没有任何指标或来历的一项变更（特殊原因）的数值之后，设置（重新设置过程控制的测量系统的基础（规则 3）
- 以上一次生产的产品为基础，通过自动补偿的方法来调整过程（规则 2）
- 在进行工作培训时，由工人 A 训练工人 B，再由工人 B 培训工人 C……，没有一个标准的培训材料。类似“邮局”的游戏。（规则 4）
- 测量了零件，并已发现偏离目标，但在控制图上的图形显示过程是受控的——因此不采取措施（规则 1）

---

<sup>12</sup>W. Edwards Deming, 《走出危机》麻省理工学院, 1982

## 第一章 — 第 C 节 测量的策略和计划

### 引言

在设计和采购测量仪器或系统之前，进行计划是很关键的。在计划阶段所做的许多决定会影响测量设备的倾向和选择。测量的目的是什么？测量结果如何使用？计划阶段将解决这些问题并且对测量过程的操作产生显著的影响，同时可以减少可能出现的问题以及以后的测量误差。

有些情况下，由于被测量零部件所涉及的风险或因为测量装置的成本和复杂性，（OEM）顾客可能要求使用 APQP 过程和承诺以决定对供方的测量策略。

不是所有产品和过程特性都要求用如此复杂的方法来开发测量系统。诸如千分尺或游标卡之类的简单标准测量工具可能不需要如此深奥的策略和计划。一个基本的原则是这零部件或次总成系统的被测特性是否已在控制计划中注明，或者是否是对产品或过程的接受决定的一个重要特性。另外，还可根据对某特定尺寸的指定公差等级来决定。在任何情况下，常识就是指导方针。

### 复杂程度

一测量系统的类型、复杂程度和目的可能驱使不同程度的项目管理、策略计划、测量系统分析，或其他测量系统选择、评估和控制的其它特殊考虑。简单的测量工具和装置（例如，天平、卷尺、固定限制或计数型量具）也许不需要更复杂的或重要的测量系统所要求的管理、计划或分析等级（例如：基准或参考标准、CMM、试验标准、自动化的线上量具等）。任何测量系统需要进行更多或少的策略计划和详细程度，取决于产品和过程情况，适当程度的决定必须由指派对测量过程和顾客负责的 APQP 小组来确定。以后续许多活动中涉及或实施的实际程度，应该由特定的测量系统、支持量具控制和校准系统的考虑、对过程深切地了解和常识来决定。

### 识别测量过程的目的

第一步是要建立测量的目的和如何应用测量。为了完成这工作，关键的是在测量过程开发之前组织一个跨功能的小组，在审核、过程控制、产品和过程开发与“测量生命周期”分析有关的特别的考量。

#### 测量生命周期

“测量生命周期（measurement life cycle）”这一概念是指随着时间的不同，对过程的了解以及过程的改进，测量方法的可能改变。例如，为了建立稳定和有能力的过程，可能开始对一个产品的特性进行测量；通过测量了解了对直接影响产品特性的关键过程控制特性。这了解意味着对产品特性信息的依靠减少了，并且可能减少抽样计划（从 5 件/小时，减少到 1 件/每班）；同时，测量的方法也可以从 CMM 测量变为某种计数型量具测量。最后可能发现可能只需要监测极少数的零件，只要过程被维护或测量和监控这维护和工装，也许就是所有必要的工作了。测量的程度是依赖着对过程理解的程度。

对供方进料，大部分的测量和监控最终可能会取消。经过相当长的时间后，在过程的同一区

域对相同的特性进行同一测量，这是对测量过程缺乏学习或停滞的证据

## 测量过程设计的选用准则

在采购一个测量系统之前，一个详细的测量过程的工程概念要被建立。利用上述所开发的目的，为了符合设计要求，各个跨功能小组将开发测量系统的计划与概念。下面是一些指南：

小组需要评价子系统或零件的设计并识别重要特性。重要特性的识别应以顾客的要求和子系统或零件在整个系统中的功能为基础。如果重要的尺寸早已被识别，则评估测量该特性的能力；例如，如果一个塑料射出零件的重要特性是它的分模线，对这尺寸的检查会很困难，测量的变差也会很高。

可用 FMEA 过程来解决上述类似的问题，即从测量零件的能力以及量具的功能两方面来分析量具设计的风险区域（设计和过程 FMEA）。这种方法同样可以用来帮助开发维护保养和校准计划。

画出过程流程图，以标示出零件或子系统的制造或装配过程中的关键过程步骤，识别该过程每个步骤中关键的输入和输出；这样有助于开发那些受过程场所影响的测量设备准则和要求。通过这样的调查，<sup>13</sup>可以得到测量计划及测量类型的清单。

对于复杂的测量系统，要作成测量过程的流程图；包括被测零件或子系统的交付、测量本身，以及退回零件或子系统到过程。

接下来使用一些头脑风暴法来开发用于每个测量所要求的共通准则。使用因果图是一种简单的方法。<sup>14</sup>参见图 2 的范例以作为思考的开始。

### 与测量计划有关的几个需要考虑的其它问题：

- 谁应加入“需求”分析中？流程图和最初的讨论将有助于关键项目的识别。
- 为什么要采取测量？及其如何被使用？数据是否用来控制、分类、资格验证等？测量的应用方式会改变测量系统的敏感等级。
- 需要什么级别的敏感度？产品规范是什么？期望的过程变差是什么？需要用量具检测的零件间的差别有多少？
- 需要对量具提供哪些信息（例如，使用手册-操作维护等），要求操作者具备哪些基本技能？谁来实施这培训？

---

<sup>13</sup> 这可以作为初始控制计划

<sup>14</sup> 参见质量控制指南，Daoru Ishikawa，亚洲生产力组织出版，1986



- 如何进行测量？是否可以进行人工测量？在移动的输送带上？离线测量？自动测量等？零件的位置和固定是否是可能的变差来源？接触或非接触测量？
- 测量如何被校准？是否将以其它的测量过程来比较？谁负责这校准的基准？
- 测量在何时何地进行？零件是否被清洁、涂油、加热等？记住，数据的使用是证实对于测量过程的一般假设，这比较保险因为这是在操作环境下收集的数据，胜过基于错误的信息下做出决定与开发一个基于非稳健环境问题下的系统。

### 研究不同的测量过程方法

在投资到新设备之前，应对目前使用的测量方法进行研究。对测量方法进行证实可以提供更可靠的操作，如有可能，使用有追踪记录证实的测量设备。

### 概念和建议方案的开发和设计

当进行概念和建议方案的开发和设计时，参见第一章第 D 节后面的“测量系统开发检查表的建议要素”。

在测量设备制造和测量过程的开发（方法、培训、文件化等）之中和之后，将进行试验研究和数据收集活动，利用这些研究和数据来理解这测量过程，从而对这过程和将来的过程进行改进。

## 第一章 — 第 D 节

### 测量资源的开发

#### 引言

本节介绍测量过程的报价 / 采购过程的寿命时间范围。已经建构了对于下列的一种自行控制的讨论：开发一测量过程报价方案的过程、获得对报价方案的回复、裁定项目、完成最终设计、开发测量过程，并在最后将已建立的测量过程与生产过程互相结合。强烈要求在没有完全阅读和理解对测量过程的所有讨论的情况下，不要使用本章节。为了从测量过程中获取最大的利益，用具有输入和输出的过程来研究并表述它。<sup>15</sup>



<sup>15</sup> 参见第一章第二节

本章内容是以团队合作的哲理所编写。这不是为购买者或采购机构所使用的工作说明。成功地完成这里所说明活动将需要小组的参与,并且应该将它纳入先期产品质量计划(APQP)小组的整体工作时间架构下管理。这样可以使不同功能的小组之间良好的合作关系—在计划过程中提出的某些概念可能会在量具供方作出最后设计之前被修改,进而满足测量系统要求。

通常,“采购过程”如于顾客和供方之间为某一特定项目所进行的正式联系。直截了当的联系是项目成功的关键,因为建立一个有效的顾客/供方关系所必需的基础工作将在这个阶段完成。顾客正式地把项目的内容列在“报价需求通知(Request For Quote,RFQ)”,即是采购过程的开始,随之而来的是供方为了符合这些意图(报价)的正式计划案说明。顾客和供方需要彻底地理解项目的要求,例如:将要交付些什么及交付方法都需要彻底地理解项目的要求,例如:将要交付些什么及交付方法都需由双方达成一致的意见。这种理解要通过双方准确及时地沟通才能达成。

一旦这概念达成了某种协定,并且已经为该项目建立了顾客/供方关系,就可以开始着手详细地设计、测量过程的制造、以及开发活动。顾客与供方的沟通在这一阶段尤其重要,因为可能有几个概念阶段要进行批准、可能的环境变化或小组成员改变,测量过程项目可能会动摇甚至失败。如果在顾客和供方之间保持并记录了经常的、详细的沟通,并由双方指派正式负责人员(分别的)维护这些沟通的话,可以降低这种风险。APQP过程是上述讨论和活动形式的理想方式。

测量过程已经被概略的设计完成之后,便可以开始测量过程/系统的相关采购活动。

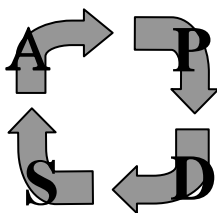
#### 数据的协调

理想上,最好使用目前普及的几何尺寸及公差(Geometric Dimensioning & Tolerancing,GD&T),在整个制造过程和测量系统中的数据需要被协调(例如,一致化)当前流行使用几何尺寸和公差(GD&T),并且应在APQP过程的非常早期建立。这项工作的初始负责人(依据特定的组织)可能是产品设计工程师、尺寸控制等。如果数据方案(datum schemes)与整个制造过程,特别是测量系统不相匹配,那么可能测量了错误的事项,可能会发生适用上的问题等,从而导致不能对制造过程实施有效的控制。

有时用于最终装配的数据方案可能与用于子部件制造过程的数据方案不匹配。如果出现了这种情况,要尽早地在APQP活动中提出来,以便小组所有的成员了解可能遇到的困难和冲突,从而尽早的利用一切机会来克服它们。在这个过程中,可能需要探索不同的数据方案以便理解这些差异的影响。

某些零件的显著特性会比其它特性产生更多的问题,例如,凸轮轴的中心,或其它圆形体、圆柱体的或管状的特性。例如,一凸轮轴必须加工成以中心线为中心,但其重要的产品特性在其凸轮。因此,制造测量时可能要求用一种演绎法中数据方案,而在最终产品测量时却要求另一种方案。

#### 先决条件及假设



在讨论开发量具供方之前,首先要假设“正确的”的产品工程产品设计(GD&T)和“正确的”过程设计(在过程中适当时间和地点进行测量)中的问题已经解决。这些假设议题不应该被转移注意,在APQP活动早期由适当的小组成员来考虑。

假设量具供方将参与APQP过程的小组活动,这量具供方对整个生产过程和产品使用有清晰的了解,所以,他的角色不仅能由他自己去了解,而且能由小组中的其它人员(制造、质量工程等)参与。

某些活动中或活动的顺序可能会有稍微的重迭，这些活动取决于特定的计划/项目或其它约束。例如，APQP 小组可能没有从量具来源取得充足的输入来开发可靠的量具概念，而其它概念可能需要量具来源的专业知识。这将因为测量系统的复杂性和小组的决定来驱使它更合理。

## 量具来源选择过程

### 开发报价方案

详细的工程概念

在正式向潜在供方提出一个正式的测量过程需求报价之前，需要开发一份测量过程的详细工程概念。小组中负责测量过程维护和持续改进的成员是开发详细工程概念的直接责任人员。他们可以是 APQP 小组的一部分。为了更好地开发这一概念，需要回答几个问题。



小组可能研究不同的议题，有助于决定在设计这测量过程时将遵循的方向和途径。有些概念可能在产品设计中就已经有规定或隐含的规定。当开发这详细概念时，需要由小组讨论的许多可能性问题可参见本节末“测量系统研究检查表建议要素”。

顾客总是过份依赖供应商的方案。在顾客要求供应商对过程问题提出建议方案之前，过程的基础和意图应由掌握过程的小组完全理解和预料。那时，只有那时过程才能被正确使用、支持和改进。

### 预防性维护的考虑

什么活动需要进行有计划的预防性维护（例如，润滑，振动分析，整体探查，更换新零件等）？大部分这些活动将取决于测量系统、设置或仪器的复杂程度。简单的量具可能只要周期性的检查，而复杂的系统可能需要持续进行详细的统计分析，并有一组工程人员通过预知性（predictive）的方式来进行维护保养

预防性（preventive）维护活动的策划应该与测量过程生恐的初衷相一致。许多活动，例如每天排干空气滤清器，在设备运行设定的若干小时后就润滑轴承等，可以在测量系统完成制造、开发和使用之前就予以规定。事实上，这样做是很完善并且能够进一步改进测量的策划与成本。与这些活动相关的数据收集方法和推荐的保养方法可以从原设备制造商处获取，或由工厂工程、制造和质量人员来开发。在测量过程完工并投入使用后，需要经常的收集与测量过程功能相关的数据并进行画图，可以使用简单的分析方法（推移图、趋势分析）用以确定系统的稳定性。最终，根据系统稳定性规定的判断，能够计划性的进行有规律的预防性维护。基于时间序列的信息对一稳定的系统进行预防性维护，将比用传统技术进行的预防性维护更经济。

## 规范 (specification)

在设计和制造过程中，规范可作为顾客和供方两者的指南。这些指南作为对可接受标准的沟通。可接受标准可能考虑了以下两个方面：

- 设计标准
- 制造标准

设计标准的形式可能取决于为项目付费的人而不同，成本问题将影响这一形式。一般来说，具有足够文件化的设计细节是一个好主意，因为这样可能由任何合格的制造商来制造或修补以符合这设计的原始意图—但是，这种决定会受到成本和关键性的限制。必要的最终设计形式可能是一些从 CAD 或工程图面上取得，这可能包括 OEM,SAE,ASTM 或其它组织所摘选的工程标准；量具供方必须取得并理解这些标准的最新版本。OEM 可能要求在启用测量系统之前进行正式的批准。

设计标准将会详细地规定将设计传送给制造者的方法（CAD—例如，CATIA、Unigraphics、IGES、手绘文件等）。对于复杂的测量系统还可能包括性能标准。

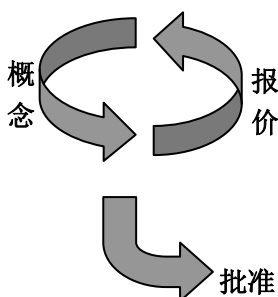
制造标准将包括测量系统制造所允许的公差。制造公差应以用于制造量具或量具零件的过程的综合能力，以及预定的被测量我的关键性为基础。制造公差不应该仅仅是产品公差的某一设定的百分数。

*如果需要生产多个夹具或系统，适当的策划和标准化可以使系统具有可互换性和便利性。*

使用标准化的零件或次总成件也使系统具有可互换性、便利性、降低成本，并且通常能减低长期测量的误差。

接到报价后，小组成员要聚集在一起对它们进行评审和评价。要注意以下项目：

### 评估报价单



- √ 是否符合基本要求？
- √ 是否有任何明显的顾虑？
- √ 是否有任何的供方提出了异常的情况，为什么？  
(异常情况可能是对价格和交付方面的重大偏离 — 没有必要把这个看成一个负面因素---一个供方可能已经发现了其它供方所忽视的地方。)
- √ 概念是否提升了简单易懂且可维修性？

---

## 可交付的文件

通常采购一测量过程有时会忽视文件作用。不能理解文件在任何成功的项目中所担任的重要意义。文件的一般作用是为了测量过程硬体在交付时，提供一套对机械的或电子的设计的原始数据（CAD 或图面）这些可以满足初始的应用要求，但是这种文件没有说明潜在的磨损部位、建议可能的故障区域，或没有说明如何使用该过程。因此用于任何过程所必要的文件要能包括测量设备的总成图和详细的零件图，及更多的文件。

用于任何系统的有效文件的作用与一张好的旅行地图的作用是一样的。例如，它可以告诉使用者如何从一个地方到另一个地方（使用者指导书或量具操作手册）。当主要路径不通或关闭时，这些文件可以告诉使用者采用其它可能的方法来完成想达到的目的地（问题诊断指南或故障诊断分析）。

一个完整的文件包可能包括：

- 一套可复制的总成和详细的机械图（CAD 或文件）  
（包括任何要求的基准）。
- 一套可复制的电子电路图、逻辑图及软件。
- 大量使用或易磨损项目/部位的建议备用零件清单。该清单应该包括可能需要较长的时间采购的项目。
- 附有机械解剖图的维修手册，以能够依步骤正确的装配或拆卸机械零组件。
- 为作业准备、操作方法及机械搬运定义了有用的要求的手册（例如，安装轴承部分）。
- 故障诊断分析与问题诊断的指南
- 认证报告（适当时可追溯到 NIST）
- 校准指导书
- 可供技术支持人员、系统操作人员和维修人员可以使用的使用者手册

上述可作为建立一份报价方案的检查清单，然而并不必包括所有内容。

这里讨论的中心议题是沟通。由于文件是沟通的一种形式，小组成员及其它人员能参与测量过程文件开发的各个阶段。

### 在供方处的鉴定

适用情况下，在测量系统供方交付之前，应对量具或测量系统进行全尺寸检验和功能测试。因此明显的，被选择的供方必须在生产现场有合格的测量设备和人员来完成以上工作。如果无法进行，必须事先安排一个合格的在外部独立实验室来完成这些工作。这种全尺寸检验和/或试验应该是基于顾客设计和制造标准要求下所执行的结果，并且应该完全记录下来以供顾客评审。

经过成功的尺寸检验后，供方应进行初期的正式的测量系统分析。同样，事先要求供方具有合格的人员、知识及经验来完成这适当的分析。顾客应该事先与供方（可能还有 OEM 参加）一起确定需要进行哪些种类的分析，并且应该了解供方可能需要的任何指导。需要进行讨论、商议或一致同意的事项包括：

- 初始 MSA 研究的目标：
  - √ 量具重复性（GR<sup>16</sup>）与重复性和再现性（GRR）的比较
  - √ 偏倚和/或线性的评估
  - √ 顾客对测量目的的评估
- 研究所需零件、测量次数和操作者数量
  - √ 接受标准
- 使用供方人员与顾客人员的比较
- 人员所需要的培训
  - √ 他们能胜任吗？
  - √ 他们理解研究的意图吗？
  - √ 可能需要使用什么软件？

此时此刻不论得到什么样结果，都应该意识到这些仅是最初分析，需要进行判断从而确定这结果是否可以接受。

---

## 运输

### 检查表

- 设备应该在什么时候交运？
- 应该如何运输？
- 谁把设备从卡车或火车上卸下来？
- 需要保险吗？
- 文件与硬件一起交运吗？
- 顾客是否有合适的设备卸下硬件？
- 交运前，系统将存放在什么地方？
- 应用后，系统将存放在什么地方？
- 交运文件是否完整并且易于被装载、搬运、卸货及安装的人员理解？

## 在顾客处的鉴定

在上述供方交运之前对测量系统所做的鉴定工作，通常在交付完成后，在顾客场所应该用同样的方式重复做一次。因为这是测量系统在其打算工作的环境下进行的首次实际研究机会，所以要慎重地考虑使用接受标准和分析方法。各团体积极参与这部分对这测量系统的最后成功以及它所产生的数据的应用极为重要。

收到测量系统后，在进行任何测量分析之前，应该对其进行全尺寸测量，从而确认它是否满足制造要求/标准。这次全尺寸检查检验工作的范围可能与测量系统在交运前在供方处所做的全尺寸结果进行比较，以获得在供方处所做的全尺寸结果的质量自信，并对在运输过程中没有损坏的依赖。当在对运输前后进行比较时，要意识到由于测量系统条件不同，两个测量结果会存在一些差异。

---

### 文件交付

为了帮助任何系统的实施或启用，至少需要以下信息：  
(这些信息要在设备交付前提供给顾客。)

- 如果小组要求，CAD 或图面文件
- 适用时，该系统的流程图
- 使用者手册
  - √ 维修/服务手册
  - √ 备用零件清单
  - √ 问题诊断指南
- 校准说明书
- 其它特别需要考虑的事项

在一开始时，已交付的文件需要被注明为“初始，这个时候可以不交付原始的或可复制的文件，因为在实施后可能需要进行修订。”事实上，直到整个系统被执行后，不提交原始文件的做法是明智的——在供方处更改文件比在顾客处更有效率。

# 测量系统开发检查清单的建议要素

本清单应该根据测量系统的情况和类型进行修改。最终检查清单的建立应该是顾客和供方合作的结果。

## 与测量系统设计和开发有关的问题：

- 被测特性是什么？** 特性的类型是什么？是机械上的特性吗？是动态的还是静态的？是一项电的特性吗？有零件内部变差大吗？
- 测量过程的结果（输出）将被应用的目的是什么？** 生产改进、生产监控、实验室研究、过程审核、出货检验、进货检验、对 D.O.E 的回应？
- 谁将使用该过程？** 操作者、工程师、技术员、检验员、审核员？
- 培训要求：** 操作者、维护人员、工程师、教室、应用实习、在职培训、学徒期间。
- 变差的原因是否已被识别？** 通过小组、头脑风暴法、渊博的过程知识，因果图或矩阵图等方法建立一个误差模型（S.W.I.P.E 或 P.I.S.M.O.E.A）
- 是否展开了测量系统的 FMEA？**
- 弹性的或专用的测量系统：** 测量系统可以是固定的、专用的还是弹性的，是否有测量不同类型零件的能力；例如：爪型量具、夹紧量具、三坐标测量机等。弹性的量具价格较贵，但从长远来看能节约成本。
- 接触式或非接触式：** 可靠性、特性类型、抽样计划、成本、维护保养、校准、人员技能要求、兼容性、环境、速度、探头的类型、零件的变形和图像处理。以上的计划可能由控制计划和测量频次确定（全面接触量具在连续抽样期间可能会过度磨损）。整个表面接触的探头、探头的类型、空气回流喷嘴、图像处理，CMM 或光学比较仪等。
- 环境：** 污垢、潮湿、湿度、温度、振动、噪声、电磁干扰（EMI）、大气流动、空气杂质等。实验室、工厂、办公室等？在小且严格的公差下以微米为测量单位的的测量系统中，以及在 CMM、光学系统、超声波仪器等环境中，环境成为一个关键的问题。对线上自动反馈类型的测量也是一个影响因素。切削油、切削碎片和极端温度也会成为问题。是否有清洁环境的要求？
- 测量及固定点：** 使用几何尺寸与公差 GD&T 清晰地定义固定搁置和夹紧点，以及在零件的何处进行测量。
- 固定方法：** 不固定或夹紧零件。
- 零件方位：** 主体的位置与其它的位置。



- 零件准备：**在测量前，零件是否应该清洁、除油、温度稳定等？
- 传感器位置：**从主定位器或定位系统的取向角度与距离。
- 相互关系问题#1— 备份的量具：**在工场内或不同工场之间是否需要备份的（或多个）量具支持？制造的考虑、测量误差的考虑、维修的考虑。标准是哪个的考虑？如何才能使每个考虑问题均符合要求？
- 相互关系问题#2— 方法差异：**在可接受的实施和操作极限内，由不同测量系统设计对同一产品/过程进行测量的测量误差结果（例如，CMM 对手动量具或开放式设定量具的测量结果）
- 自动或手动：**线上、线外、操作者依赖性。
- 破坏性的与非破坏性（NDT）的测量：**例如：拉力试验、盐雾试验、电镀/涂装厚度、硬度、尺寸测量、图像处理、化学分析、应力、耐久性、冲击、扭力、扭矩、焊接强度、电特性等。
- 潜在测量量程：**可能的测量尺寸大小和期望的量程。
- 有效的分辨率：**对应用在一特殊的应用场所的可接受性，如测量对物理变化是否敏感（探测过程或产品变差的能力）？
- 灵敏度：**最小的输入量的信号能产成一个可探测的输出（可辨别的）信号，测量装置对应用这种情况的可接受性？灵敏度由量具的固有设计和质量（OEM）、使用期间的维护和操作条件所决定。

#### 与测量系统制造问题（设备、标准、仪器）：

- 是否已在系统设计中针对变差源的识别？**设计评审、验证和确认。
- 校准和控制系统：**推荐的校准计划及设备审核及其文件。频率、内部的或外部的、参数、过程中验证检查。
- 输入要求：**机械的、电子的、液压的、真空的、波动抑制器、干燥器、滤清器，准备作业和操作问题、隔离、解析度和灵敏度。
- 输出要求：**类比或数位、文件和记录、档案、保存、存取、备份。
- 成本：**开发、采购、安装、操作和培训的预算要素。
- 预防性维护：**形式、计划、成本、人员、培训、文件。

- 可维修性：**内部的和外部的、场所、支持程度、回应时间、服务配件的可获得性、标准零件清单。
- 人机工程学：**在长时间装载和操作设备过程中，人员不被伤害的能力。测量装置的讨论需要着重在测量系统与操作者之间相互关系。
- 安全考虑：**人员、操作、环境、切断。
- 贮存和场所：**建立对测量设备的贮存及场所的要求。隔离、环境、安全、获得性（接近）有关的问题。
- 测量周期时间：**测量一个零件或特性需要多长时间？测量周期要与过程和产品控制合并。
- 是否有任何对工艺流程、批量完整性、记录、测量和零件回复的干扰？**
- 材料搬运：**需要特殊架子、支撑夹具、运输设备或其它物料搬运设备来放置被测量的零件或测量系统本身？
- 环境问题：**是否有特殊环境要求、条件、限制等影响本测量过程或临近的过程？是否要求特殊排气？有温度和湿度控制的必要吗？湿度、振动、噪声、电磁干扰、清洁？
- 有特殊的可靠性要求或考虑吗？**设备是否能在任何时间下维持其状况？在生产使用前是否需要进行验证？
- 备用零件：**共享清单、适当的供应和订购系统、可获得性、导入期的理解与说明。是否有足够的安全库存？（轴承、软管、皮带、开头、插座、阀等。）
- 使用者说明：**夹紧顺序、清洁程序、数据解释、图表、目视辅具、易于理解的。可获得性、适当的陈列
- 文件：**工程图面、诊断分析、使用者手册、语言等。
- 校准：**可接受的标准进行比较。可接受的标准的可获得性和费用。建议的频率、培训要求、停机时间的要求
- 贮存：**是否有与测量装置贮存有关的特别要求或考虑？隔离、环境、防止损坏/窃盗等。
- 防错：**已知的测量程序错误？是否容易由操作者更正（非常容易？）数据输入、设备误用、防错。

**与测量系统实施有关的问题（过程）：**

- 支持：**由谁支持测量过程？实验室技术人员、工程师、生产、维修保养、与外部签订保

养合同？

- 培训：** 为了使用和维护本测量过程，需要为操作者/检验人员/技术人员/工程师提出哪些培训？时间、资源和费用。谁来培训？在哪进行培训？导入期要求？与测量过程的实际使用相协调。
- 数据管理：** 如何管理从本测量过程输出的数据？人工、电脑化、汇总法、汇总频率、评审方法、评审频率、顾客要求、内部要求。可获得性、存储、存取、备份、安全、数据解释。
- 人员：** 是否需要聘请人员以支持这一测量系统？成本、时间、可获得性有关的问题。目前的或新的。
- 改进方法：** 有谁来对测量过程进行经常性的改进？工程师、生产人员、维护保养人员、质量人员？使用什么样的评价方法？是否有一系列以识别需要的改进？
- 长期稳定性：** 长期性研究的评估方法、形式、频率、需求。漂移、磨损、污染、操作完整性。长期误差是否能被测量、控制、理解和预测？
- 特别的考虑：** 检查人员的特质、体能限制或健康问题：色盲、视力、身体强度、疲劳、耐力、人机工程。

## 第一章— 第五节 测量问题

### 引言

在评价一个测量系统时必须考虑三个基本问题：

1) 测量系统必须证明有足够的敏感度。

√ 首先，仪器（和标准）是否有足够的分辨力？选择测量系统的基本开始点是由设计所决定的分辨力（或等级）。通常应用 10: 1 的原则，也就是说，仪器的分辨力应该把公差（或过程变差）细分为 10 分之一或更多。

√ 第二，测量系统是否证明具有有效的解析度？与分辨力类似关系，有效解析度要确定测量系统的应用状况下具有敏感度以检测出产品或过程变差。

2) 测量系统必须稳定。

√ 在重复性的状况下，测量系统变差仅由普通原因而不是由特殊原因（无秩序的）产生。

√ 测量分析必须时时刻刻考虑实践和统计的显著性。

3) 统计的特性（误差）要一直保持在期望的范围内，并且足以满足测量的目的（产品控制或过程控制）。

长期以来，传统上只用公差的百分比作为测量误差的报告方法，这已不能适应于过程持续改进的市场竞争的策略。随着过程改变和改进，测量系统必须为它的使用意图重新评价。组织（管理者、测量计划者、生产操作者和质量分析人员）理解测量的目的并进行适当的评价是不可或缺的。

### 测量系统变差的类型

测量通常被假定是精确的，而且许多后续的分析及结论都是以该假设为基础。个人可能无法意识到测量系统中存在影响各个测量的变差，它进而又影响基于这些数据的结论。测量系统误差可以分成五种类型：偏倚、重复性、再现性、稳定性和线性。

测量系统研究的目的之一是获得测量系统与其所处环境有相互作用时，其产生的测量变差的类型和结果的讯息。这些 信息是很有价值的，因为对一般的生产过程，确认重复性和校准偏倚，以及为这些制定合理的极限，比提供具有非常高重复性的准确量具更实用。应用这种研究可提供：

- 接受新测量设备的准则。
- 一个测量装置与另一个测量装置的比较。
- 评价一个疑似不充分量具装置的依据。
- 测量设备维修前后的比较。
- 为计算过程变差以及生产过程可接受程度的必要构成元素。

- 绘制量具性能曲线（GPC）<sup>17</sup>的必要信息，GPC 表示接受某一零件真值的概率。

下列定义帮助我们描述与一测量系统有关的误差或变差类型，以便在其后讨论中可以清楚地理解每个术语。每种定义都给出了说明，以便生动地展现每项术语的含义。

## 变差的定义及潜在的来源

### 可操作的定义

“可操作的定义系指人员可以进行其业务的定义。一个安全的、圆的、可靠的或其它任何质量（特性）的可操作性的定义必须是可相互交流的，对于卖主和买主具有同样的意义，对于生产工人来说，在今天和昨天都具有同样的意义。例如：

- 1.一块材料或一个总成件的特定试验
- 2.判断用的准则
- 3.决定：是或不是，物体或材料符合或不符合准则<sup>18</sup>

## 标准（standard）

标准通常指全体同意作为一个比较的基础，一种可接受的形式。它可以是由权威机构设定或建立的一个物品或是一个整体（仪器、程序等），如用来度量数量、重量、长度、值或质量的规则。

这整体（ensemble）的概念是在美国国家标准协会（ANSI）/美国质量控制协会（ASQC）标准 M1-199619 中正式形成。这术语强调需要考虑所有影响测量不确定度的实情；如环境、程序、人员等。“举一个简单的整体例子，校准块规的整体包括了标准块规、比较测定仪、操作者、环境及校准程序”。

## 参考标准（reference standard）

指在某一特定地点存放的，通常具有最高度量衡的质量的标准。使用这样的标准所做的测量是可以追溯的。

## 测量和试验设备（M&TE）

完成一次测量所需要的所有测量仪器、测量标准，基准材料以及辅助设备。

---

<sup>17</sup> 见第五章，第三节

<sup>18</sup> W.E.戴明,走出危机(1982,1986),P277。

<sup>19</sup> 这个定义被后来的军方标准修订为测量和试验设备或 M&TE。

### 校准标准 (calibration standard)

进行日常校准时用来作为一项参考的标准。意图是作为校准工作量和实验室参考标准之间的缓冲。

### 传递标准 (transfer standard)

用来将某一已知数值的标准与即将进行校准的单元进行个别比较的标准。

### 基准 (master)

作为在校准过程中的一项标准。也可以被描述为参考或校准的标准。

### 工作标准 (working standard)

在实验室内进行日常测量的标准。其不作为校准标准使用，但是也许可以用作传递标准。

为物件选择一个标准应该很慎重。所采用的标准应能够反映该测量系统的用途和范围，以及以时间为基础的变差，如磨损及环境因素（温度，湿度等）。

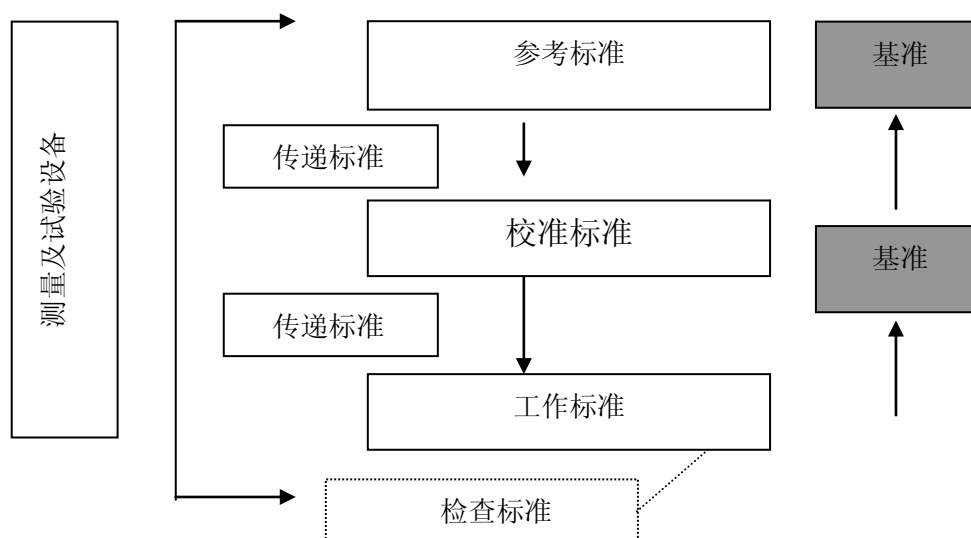


图 3：不同标准之间的联系

### 检查标准 (check standard)

一个与被指定的测量过程非常相似的测量物品，但它比被评价的测量过程更稳定。

### 参考值 (reference value)

参考值也称为可被接受的参考值或基准值。它是一个被同意用作参考比较的一个物品或整体的值。可接受的参考值取决于以下几个方面：

- 用高一级的测量设备（如度量衡实验室或全尺寸检验设备）多次测量平均值确定。
- 法定值：由法律定义并强制执行的。
- 理论值：以科学原理为基础。
- 指定值：以一些国家或国际组织的实验工作为基础（由健全的理论支持）。
- 一致同意的值：以一些科学或工程组织赞助的合作实验为基础，通过专业人士和商业组织等使用者一致同意其定义的值。
- 协议值：由受影响的各团体所协调一致得到的值。

在所有情况下，参考值需要以一个可操作的定义为基础。并且由一个可接受的测量系统得到其结果。为了达到这目标，用来确定这参考值的测量系统应该包括：

- 测量仪器与正常评价的系统相比，具有较高的分辨力和较低的测量系统误差。
- 可追溯到（美国）国家标准和技术局（NIST）或其他的 NMI 的标准进行校准。

### 真值 (true value)

真值是零件的“实际值”，尽管该值不被知道且无法知道，但它是测量系统的目标。所有个别的值都要尽可能地（经济地）接近这个值。遗憾的是，永远不可能知道确切的真值。在所有的分析中，参考值被作为真值的最佳近似值。由于参考值被用作替代真值，所以这些标准术语常常互换使用，不过不推荐这种用法。<sup>21</sup>

### 分辨力

分辨力是仪器可以探测到并如实显示的参考值的变化量。通常也被称为可读程度或解析度。

这种测量的能力通常是某测量仪器上最小刻度的值。如果仪器刻度较“粗略”，则可以使用刻度值的一半作为解析度。

图 4: 分辨力

<sup>21</sup> 同样见 ASTM: E177-90a。

可以将上述的比例规则视为确定分辨力的一个起始点，因为这还没有包括测量系统变差的任何其它因素。

由于经济和物理上的限制，测量系统不能识别过程分布中成各个独立或不同测量特性的所有零件。取而代之的是是以这些测量特性分组成不同的数据种类。在同一数据分类中的所有零件，对其被测特性来说具有相似的数值。

如果测量系统的分辨力（灵敏度或有效解析度）不足，它将不是一个识别过程变差的合适的系统，或对各个零件我值定量。如果是这样的情形，应该使用更好的测量技术。

如果不能探测过程变差，则这分辨力用于分析是不可接受的；如果不能探测出特殊原因的变差，它用于控制是不可接受的（见图 5）。

分类数	控制	分析
1 组数据分类	<p>只有在下列条件下才可用于控制：</p> <ul style="list-style-type: none"> <li>● 与规范相比，过程变差较小</li> <li>● 在预期过程变差上，损失函数很平缓</li> <li>● 过程变差的主要来源导致平均值偏移</li> </ul>	<ul style="list-style-type: none"> <li>● 用于过程参数和指数的估计是不可接受的</li> <li>● 只能指出过程是否在产生合格或不合格的零件</li> </ul>
2—4 个数据分组	<ul style="list-style-type: none"> <li>● 根据过程分布，可用于半计量控制技术</li> <li>● 可以产生不敏感的计量控制图</li> </ul>	<ul style="list-style-type: none"> <li>● 用于过程参数及指数的估计一般是不可接受的，因为它仅提供粗劣的估计</li> </ul>
5 个或更多的数据分组	<ul style="list-style-type: none"> <li>● 可用于计量控制图</li> </ul>	<ul style="list-style-type: none"> <li>● 建议使用</li> </ul>

**图 5：过程分布的区别分类数（ndc）对控制与分析活动的影响**

不适用的分辨力可能会在极差图中出现。图 6 是包含了从相同数据中取得的两组控制图。控制图（a）显示精确到千分之一英寸的原始测量值，控制图（b）表示这些百分之一英寸的原始测量值是四舍五入的。由于人为的紧缩限制，控制图（b）显示是失控的。这四舍五入的零极差超越了描述这一产品的子组变差的象征。

不合适的分辨力的最佳象征，可由过程变差的 SPC 极差图看出来。尤其是，是当极差图显示可能只有一个，二个或三个极差值 在控制限内时，这种测量就是在分辨力不足时进行的。另外，如果极差图显示有四种可能的极差值在控制限内，并且超过四分之一以上的极差值为零，则该测量也是在分辨力不足下进行的。另一个不合适的分辨力是一个普通的概率图，其数据将被堆叠，而不是沿着 45 度线流动返回到图 6，控制图（b）可能仅有两个极差值在控制限之内（值 0.00 和 0.01）。因此，上述规则正确地识别出失控的原因是分辨力（灵敏度或有效分辨率）不足。



当然，可以通过提高测量分辨力，从而改变检定子组内变差的能力来补救这一问题。如果相对于过程变差，测量系统的可视解析度较小，那么这个测量系统就有足够的分辨力。因此为得到足够的分辨力，建议解析度最多可以到全过程的  $6\sigma$  标准偏差的十分之一，以取代传统上设定的规则，即可视解析最大为公差范围的十分之一。

最终情况是，能获得一个稳定的、高能力的过程，且该过程使用了一稳定的、在实施技术限制内最好等级的测量系统。有效解析度可能不够，那么对测量系统进行进一步的改进便成为不切实际的。在这些特别情况下，测量计划可能需要使用其它替代的过程监控技术。只有具有一定资格的，熟悉测量系统和过程的技术人员，才能作出决定并予以文件化。这种情况要求必须获得顾客的批准，并在控制计划中予以规定。

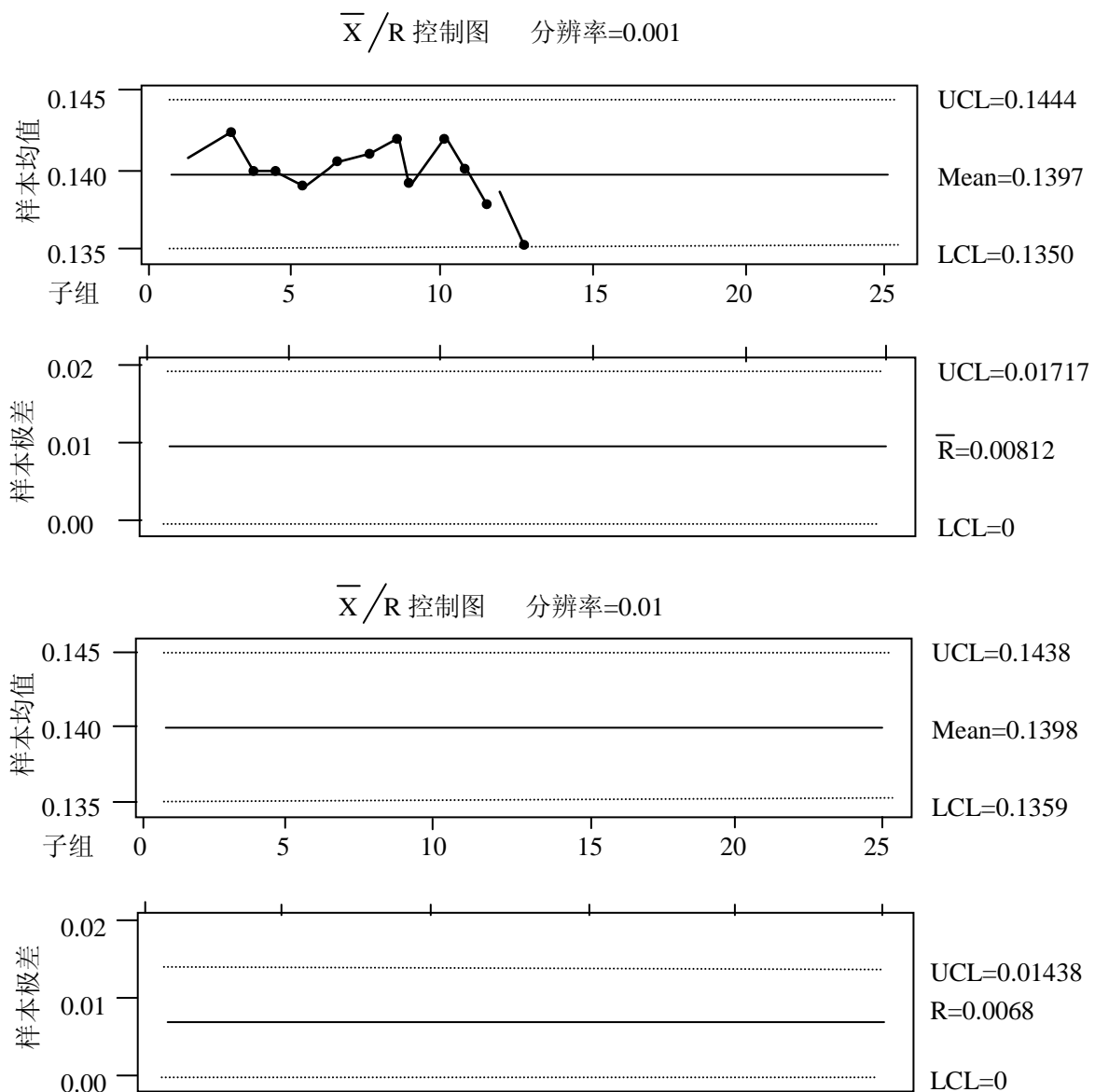


图 6: 过程控制图<sup>22</sup>

<sup>22</sup>图 6 摘自《Evaluating The Measurement Process》,由 Wheeler 和 Lyday 编著。

## 测量过程变差

对大多数测量过程，其所有的测量变差通常被描述为正态分布。正态概率是对一个测量系统分析的标准方法的假设。事实上，有一些测量系统不是正态分布，当这些情况发生时，而却以正态分布为假设，则 MSA 方法可能会过大地估计评价测量系统误差。因此，这测量分析必须认清该非正态分布的测量系统，并修改评价方法。

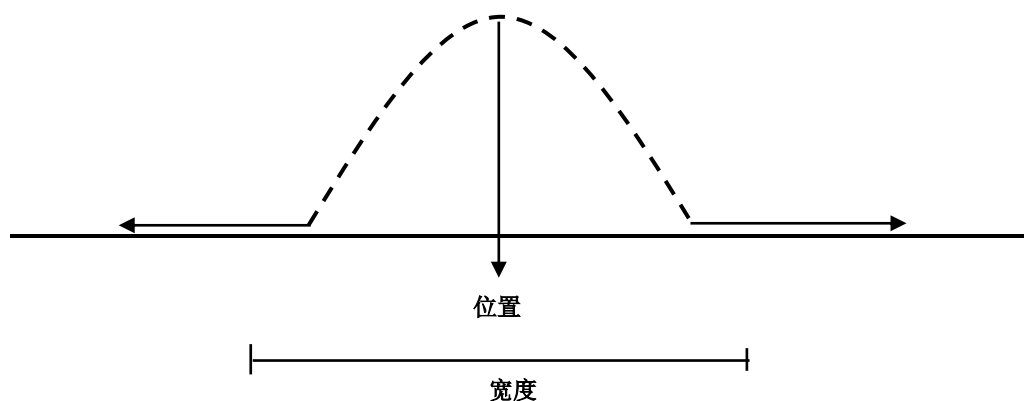


图 7：测量过程变差的特性

## 位置变差

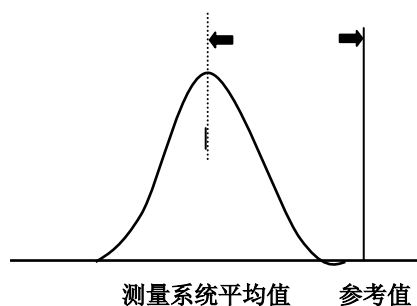
### 准确度 (Accuracy)

通用的正确说法，它是指一个或多个测量结果的平均值与一个参考值之间一致的接近程度。测量过程必须处于统计控制状态，否则过程的准确度就毫无意义。

在某些组织内，准确度通常与偏倚互用。ISO（国际标准化组织）和 ASTM（美国实验与材料协会）使用准确度这个术语来涵盖偏倚和重复性。为了避免准确度这一术语的混淆使用，ASTM 建议只用偏倚来描述位置误差。接下来的内容也将遵循这一原则。

### 偏倚

偏倚通常被称作“准确度”。因为“准确度”有多种意思，所以建议不要用准确度来替代“偏倚”。



偏倚是对相同零件上同一特性的观测平均值与真值（参考值）的差异。

译者注，原著中没有第 23 注释。

偏倚是测量系统的系统误差。它会增进所有已知的或未知的变差来源所共同影响的总偏差，这促使在某一测量时期内重复的应用相同测量过程时，以总偏差趋向去恒定和预测地补偿所有的结果。

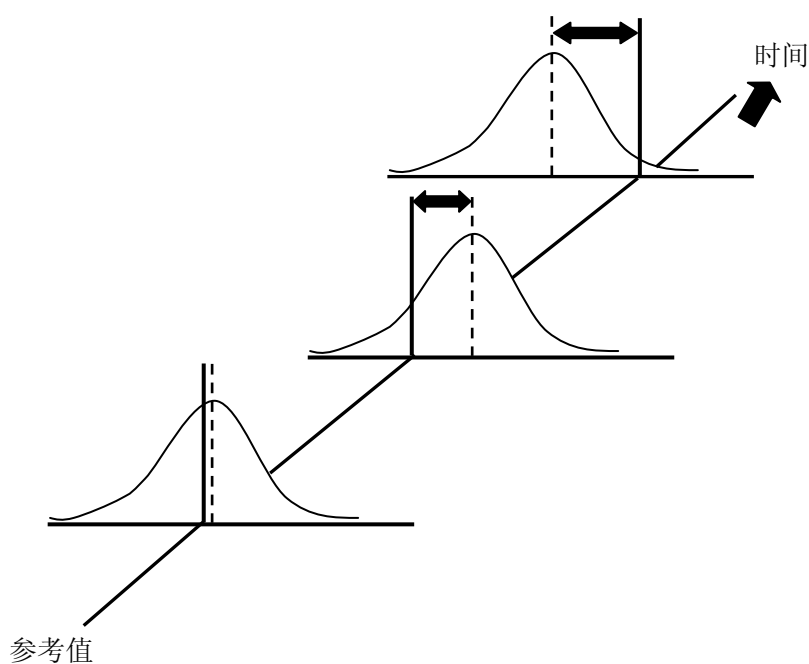
造成过大偏倚的可能原因有：

- 仪器需要校准
- 仪器、设备或夹具磨损
- 基准的磨损或损坏，基准偏差
- 不适当的校准或使用基准设定
- 仪器质量不良——设计或符合性
- 线性误差
- 应用错误的量具
- 不同的测量方法——作业准备、载入、夹紧、技巧
- 测量的特性不对
- 变形（量具或零件）
- 环境——温度、湿度、振动、清洁
- 错误的假设，应用的常数不对
- 应用——零件数量、位置、操作者技能、疲劳、观察误差（易读性、视差）

在校准过程中所使用的测量程序（如：使用“基准”）应该尽可能地与正常操作的测量程序一致。

### 稳定性 (stability)

稳定性（或漂移）是指经过一段长期时间下，用相同的测量系统对同一基准或零件的同一特性进行测量所获得的总变差。也就是说，稳定性是整个时间的偏倚变化。



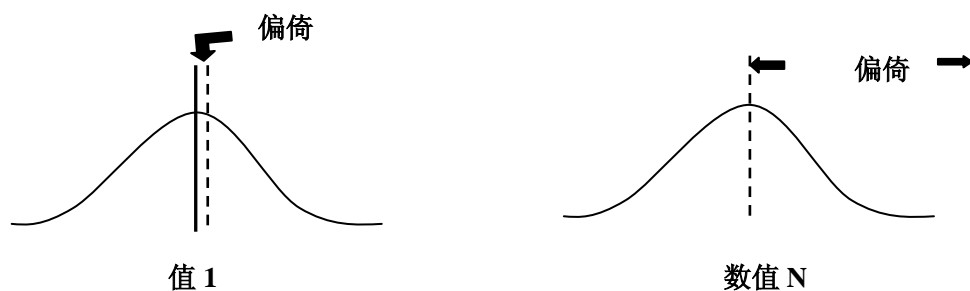
造成不稳定性的可能因素：

- 仪器需要校准，缩短校准周期
- 仪器、设备或夹具的磨损
- 正常老化或损坏
- 维护保养不好——空气、动力、液体、过滤器、腐蚀、尘土、清洁
- 基准的磨损或损坏，基准的误差
- 不适当的校准或使用基准设定
- 仪器质量不好——设计或符合性
- 仪器缺少稳健的设计或方法
- 不同的测量方法——作业标准、载入、夹紧、技巧
- (量具或零件) 变形
- 环境变化——温度、湿度、振动、清洁
- 错误的假设，应用的常数不对
- 应用——零件数量、位置、操作者技能、疲劳、观察错误（易读性、视差）

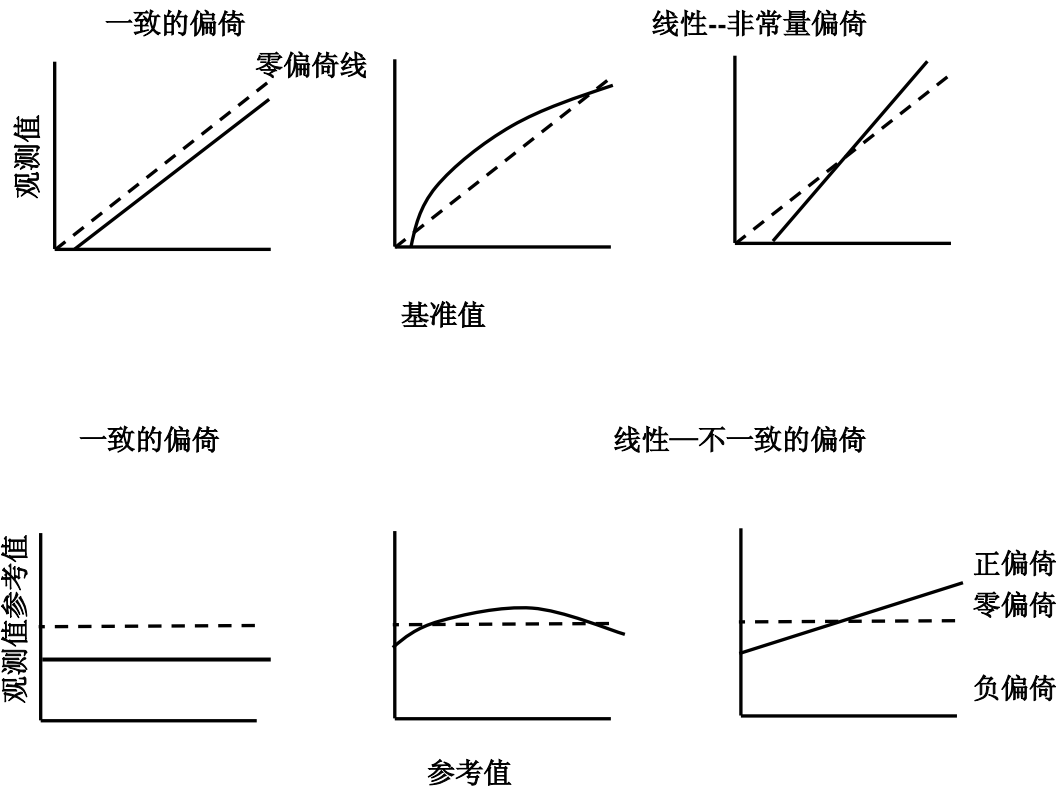
### 线性 (linearity)

线性是在设备预期的工作 (测量) 量程内，偏倚值的差异。线性可被视为对于量程大小不同所发生的变化。

注意不可接受的线性可能以各种形式出现。不要假定一个常量偏倚。



注意:不可接受的线性导致不同的情况。不要假定偏倚是不变的。



造成线性误差的可能原因如下：

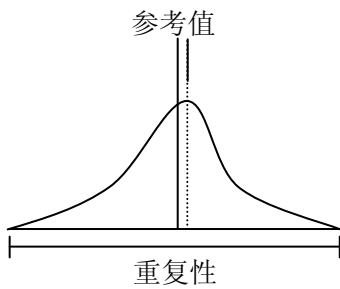
- 仪器需要校准，缩短校准周期
- 仪器、设备或夹具的磨损
- 维修保养不好——空气、动力、液体、过滤器、腐蚀、尘土、清洁
- 基准的磨损或损坏，基准的误差——最小/最大
- 不适当的校准（没有涵盖操作范围）或使用基准设定
- 仪器质量不好——设计或符合性
- 缺乏稳健的仪器设计或方法
- 应用了错误的量具
- 不同的测量方法——作业准备、载入、夹紧、技巧
- 随着测量尺寸不同，（量具或零件）变形量不同
- 环境变化——温度、湿度、振动、清洁
- 错误的假定，应用的常数不对
- 应用——零件数量、位置、操作者技能、疲劳、观察误差（易读性、视差）

### 宽度变差

精密度（precision）

传统上，将测量系统整个作业量程范围内（尺寸、范围和时间）的解析度、敏感度和重复性的最终影响定义为精确度。在一些组织内，精确度和重复性互换使用。事实上，精确度最常用在描述测量范围内重复测量的预期变差，测量范围可能是尺寸或时间（例如“某测量设备在高量程和低量程一样精确”或者“今天和昨天一样精确”）。精密度也是对应重复性的线性

偏倚量（虽然重复性是随机的，而偏倚是系统的误差）。ASTM（美国实验及材料协会）定义的精确度的范围更广，包括由于不同读数、量具、人、试验或条件造成的变差。



### 重复性 (repeatability)

传统上把重复性称为“评价人内变异”。重复性是用一个评价人使用相同的测量仪器对同一零件上的同一特性，进行多次测量所得到的测量变差；它是设备本身的固有变差或能力。重复性通常被称为设备变差，但这是一种误解。事实上，重复性是指在指定的测量条件下连续测量的普通原因（随机误差）的变差。重复性的最佳措辞为：当测量条件已被确定和定义——以确定的零件、仪器、标准、方法、操作者、环境和假设之下，系统内部的变差。除了设备内部变差以外，重复性还包括在误差模型中的任何条件下的内部变差（见下文）。

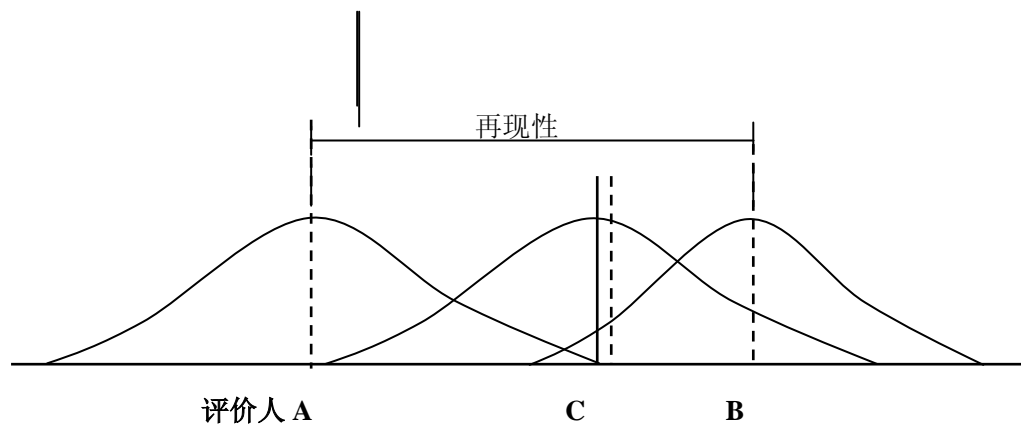
重复性不好的可能原因包括：

- 零件内部（抽样样本）：形状、位置、表面光度、锥度、样本一致性
- 仪器内部：修理、磨损、设备或夹具的失效、质量或保养不好
- 标准内部：质量、等级、磨损
- 方法内部：作业准备、技术、归零、固定、夹持、点密度的变差
- 评价人内部：技巧、位置、缺乏经验、操作技能或培训、意识、疲劳
- 环境内部：对温度、湿度、振动、清洁的小幅度波动
- 错误的假设，稳定，适当的操作
- 仪器设计或方法缺乏稳健性，一致性不好
- 量具误的用
- （量具或零件）失真，缺乏坚固性
- 应用——零件数量、观测误差（易读性、视差）

### 再现性 (reproducibility)

传统上把再现性看作“评价人之间”的变差。通常将再现性定义为由不同的评价人，采用相同的测量仪器，对同一零件的同一特性进行测量所得的平均值的变差。手动仪器通常的确受操作者技能的影响，但是，对于操作者不是变差主要原因的测量过程（即自动的系统），上述说法是不正确的。因为这原因，再现性是指测量的系统之间或条件之间的平均值变差。

ASTM（美国实验及材料协会）的定义比上述定义更广，不仅包括评价人不同，同时可能还包括：量具、实验室及环境（温度、湿度）的不同，除此之外，在再现性计算中还包括重复性。



### 造成再现性误差的潜在原因包括：

- 零件（抽样样本）之间：使用相同的仪器、操作者和方法测量 A、B、C 零件类型时的平均差异。
- 仪器之间：同样零件、操作者和环境下使用 A、B、C 仪器测量的平均值差异。注意：在这种情况下，再现性误差通常还混有方法和/或操作者的误差。
- 标准之间：测量过程中，不同的设定标准的平均影响。
- 方法之间：由于改变测量点密度，手动或自动系统、归零、固定或夹紧方法等所造成的平均值差异。
- 评价人（操作者）之间：评价人 A、B、C 之间由于培训、技巧、技能和经验所造成的平均值差异。推荐在为产品和过程鉴定和使用手动测量仪器时使用这种研究方法。
- 环境之间：在第 1、2、3 等时段所进行的测量，由环境周期所造成的平均值差异。这种研究常用在使用高度自动化测量系统对产品和过程的鉴定。
- 研究中的假设有误
- 缺乏稳健的仪器设计或方法
- 操作者训练的有效性
- 应用——零件数量、位置、观察误差（易读性、视差）

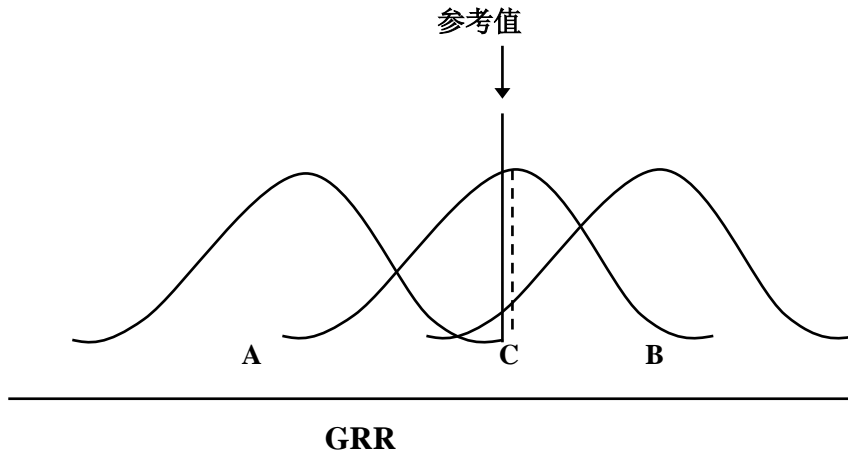
正如上述两种定义中介绍的那样，ASTM 所使用的定义和本手册使用的定义有区别。ASTM 文献专注的是对室内实验室评价，其集中在实验室与实验室的差别，包括可能不同的操作者、量具和环境以及实验室内部的重复性的影响。因此，他们定义需要包含这些差异。按照 ASTM 标准，重复性是最好的设备在当前状态下（一名操作者，一个量具，短时间内）的差异，而再现性是当差异原因来自于多个来源时，呈现更多典型操作状况。

### 量具 R&R 或 GRR

量具的 R&R 是结合了重复性和再现性变差的估计值。换句话说，GRR 等于系统内部变差和

系统之间变差的和。

$$\sigma^2_{GRR} = \sigma^2_{\text{再现性}} + \sigma^2_{\text{重复性}}$$



### 敏感度

敏感度是指能产生一个可检测到的（有用的）输出信号的最小的输入。它是测量系统对被测特性变化的回应。敏感度由量具设计（分辨力）、固有质量（OEM）、使用中保养以及仪器操作条件和标准来确定。它通常被表示为一测量单位。

影响灵敏度的因素包括：

- 一个仪器的衰减能力
- 操作者的技能
- 测量装置的重复性
- 对于电子或气动量具，提供无漂移操作的能力
- 仪器使用所处的环境，如大气条件、尘土、湿度

### 一致性

一致性是系统随时间变化，测量变差的差值。它可以看成是重复性随时间变化的差值。

影响一致性的因素都是特殊原因变差，包括：

- 零件的温度



- 电子设备必要的预热
- 设备磨损

### 均一性

均一性是量具整个工作量程内变差的差值。它可以看成是在不同尺寸值下重复性的同质性（相同性）。

影响均一性的因素包括：

- 由于位置不同，夹具能允许更小/更大的尺寸
- 刻度的可读性不够
- 读数的视差

### 测量系统变差

#### 能力

一测量系统的能力是基于在短期评估中测量系统误差（随机的和系统的）的组合变差的估计值。简单的能力包括下列组成要素：

- 不准确的偏倚或线性
- 重复性或再现性（GRR），包括短期的一致性

参考第三章对以上典型要素计算的方法和范例。

因此，一测量能力的评估值，表示某测量系统在指定条件、范围和工作量程下的预期误差（不同于测量不确定度，其表示与测量结果有关的误差或数值的预期范围）。能力是一个组合变差，当测量误差互不相关时（随机并且是独立的）的能力可由下公式计算得到：

$$\sigma^2_{\text{能力}} = \sigma^2_{\text{偏倚(线性)}} + \sigma^2_{\text{GRR}}$$

理解和正确地应用测量能力要记住以下两点：

首先，能力的估计总是与设定测量——条件、量程和时间的范围有关。例如，没有限定测量条件的范围和或工作量程，而描述一个 25mm 毫米尺的能力是 0.1mm 是不完整的。再者，这也是为什么一个误差模型对于定义测量过程是如此重要的原因。一估计测量能力的范围，要在有限的部分测量量程范围内或整个的测量量程内，对操作非常确切的或是一般的陈述。短期的意思是：对一系列测量周期的能力；完成 GRR 评价的时间；一特定的生产期间，或者以校准频率的时间表示。一测量能力的表达要尽可能完整，从而能够合理的呈现出测量的条件和范围。可以用文件化的控制计划来说明这意图。

第二，在一个能力估计中要包括测量工作量程范围内的短期一致性和均匀性（重复性误差）对于一个简单的量具，如一把 25mm 的分厘卡，用代表性的、熟练的操作人进行的量具整个量程内的测量重复性，可以看成是一致并且均一的。在本例中，一能力估计可能包括了在一般条件下，对量具整个测量量程内的多种特性。更大量程或更复杂的测量系统（如 CMM）

可能要证明（未修正的）线性、均一性和在整个量程或尺寸内的短期一致性等测量误差。因为这些误差是相互关联的。所以他们不能用上面简单的线性公式来组合。当（未修正的）线性、均一性或一致性在整个量程内的变化很大时，测量计划者和分析人员就只有两种实际选择：

- 1) 报告对于整个规定的条件、范围和测量系统量程的最大（最坏的情况）能力或
- 2) 确定并报告对测量量程中特定多个能力评估值（例如：低、中、高量程）

## 性能

像过程性能一样，测量系统性能是所有重大的决定性的变差来源的长期总影响。性能是组合了测量误差（随机的和系统的）的长期评估计算。因此，性能包含了以下长期误差为要素：

- 能力（短期误差）
- 稳定性和一致性

参考第三章对以上典型要素计算的方法和范例。

测量性能的估计是一个规定的条件、范围和测量系统量程的预期误差的表达（不同于测量不确定度，测量不确定度是一个与测量结果有关的误差或值的预期范围的表达）。当测量误差互不相关时（随机的和独立的），该合成变差（方差）的性能表达式可以量化为：

$$\sigma^2_{\text{性能}} = \sigma^2_{\text{能力}} + \sigma^2_{\text{稳定性}} + \sigma^2_{\text{一致性}}$$

此外，如同短期能力一样，长期性能总是与设定测量——条件、量程和时间的范围有关。一估计测量性能的范围，要在有限的部分测量量程范围内或在整个测量范围内，对操作非常确切的或是一般的陈述。“长期”意味着：一定时期内几个能力评估的平均值；由测量控制图计算出的长期误差的平均值、校准记录的评估或多个线性研究，或者是以测量系统整个生命周期和量程内多次 GRR 研究的平均误差。表达一测量性能时需要尽可能完整，从而可以合理地呈现出测量条件和量程。

在一性能估计中要包括整个测量量程内长期一致性和均一性（重复性误差）。测量分析人员必须意识到这些误差之间的潜在的相关性，以免过分地估计这性能。这要依赖于这些误差被怎样确定。当长期（未修正的）线性，均一性和一致性在整个量程内变化有很大差距时，测量计划者和分析者就只有以下两个实用的选择：

- 1) 报告该测量系统整个指定的条件、范围和量程中的最大性能（最坏的情况）或
- 2) 确定并报告这测量量程中特定部分的几个性能评估值（例如低、中、高量程）。

## 不确定度

由 VIM 定义的测量不确定度为“是一个与测量结果有关的参数，其特性是由于被测物特征所可能合理造成的数值离散。<sup>24</sup>更详细的内容，参见第一章，第 F 节。

---

<sup>24</sup> 被测对象当时被 VIM 定义为“用来测量的一定量的物体”。

## 评论

对于测量系统的参数、准确度和精密度，大部分的操作人员很熟悉，因为他们每天都要在技术或销售讨论中用到。但是，这些术语也很容易混淆，因为他们通常以为这些术语是可以互换的。例如，如果量具由独立的代理机构认证它是很准确的，或这设备由其供方保证它是很精确的，而这会被不正确地认为所有的读数会与真值很接近。这说法不只是理解性的错误，更可能导致对产品和过程的错误决定。

对于偏倚和重复性也有同样的混淆（如准确的和精确的测量一样）。了解以下几点是很重要的：

- 偏倚和重复性是相互独立的（见图 8）
- 控制了一些误差来源中的一个原因，不能保证控制了其它的误差原因。因此，测量系统控制计划（传统上被称为量具控制计划）要对所有相关的变差来源进行量化和跟踪。<sup>25</sup>

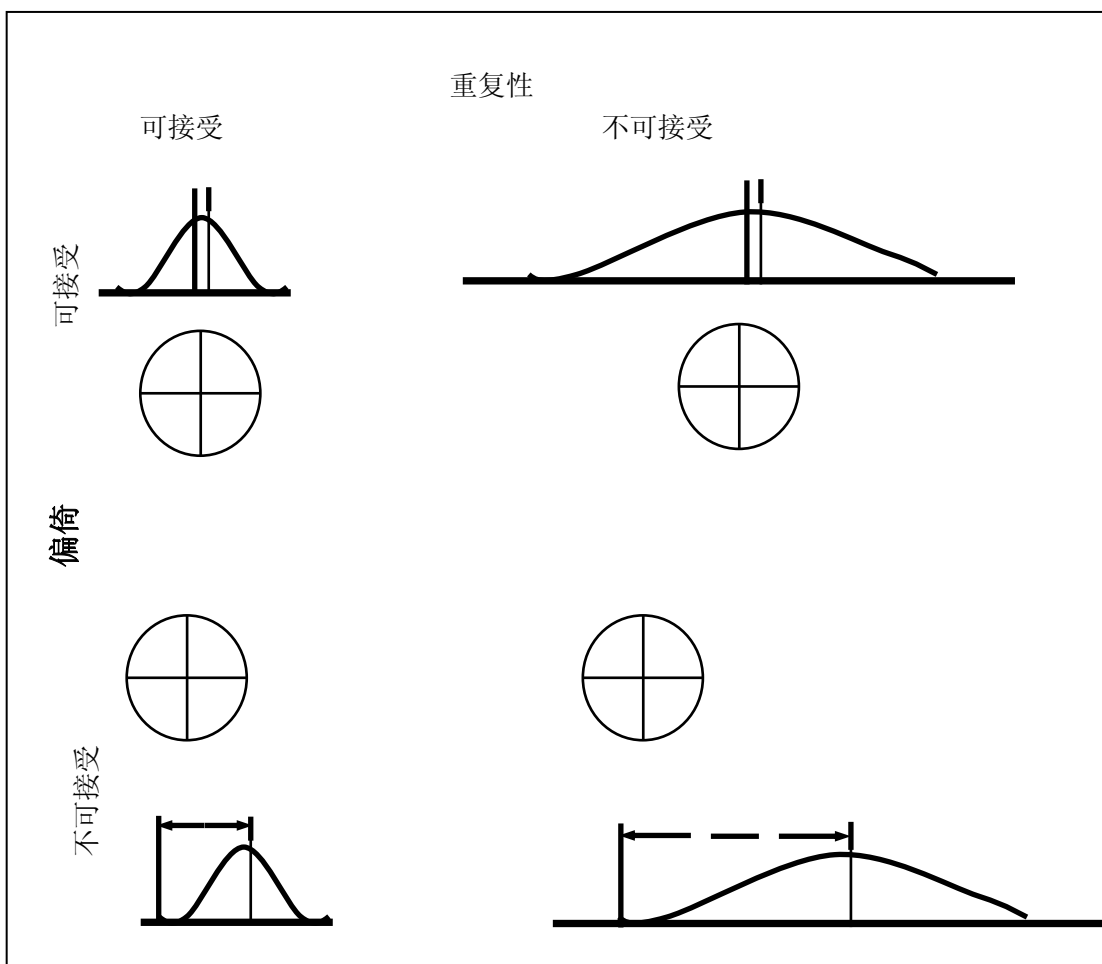


图 8：偏倚和重复性的关系

<sup>25</sup> 参见第一章，第 B 节

## 第一章—第F节 测量不确定度

---

### 总则

测量不确定度是国际上用来描述一个测量值的质量的术语。该术语传统上为许多在度量衡或量具实验室所进行的高准确度测量使用，如 QS-9000 或 ISO/TS16949 之类的质量系统标准要求：“要了解测量不确定度，并与其它任何检验、测量或试验设备所需的测量能力相一致”。<sup>26</sup>

本质上，不确定度是对一指定的测量结果范围的描述，在特定的置信度水准内，预期的包含了实际的测量结果的范围。不确定度是测量可靠性的一种量化的表达。这种概念可简单的表达为：

测量实际值=测量的观测值（结果）±U

U 是一个被测量物和测量结果的“扩展不确定度”。扩展的不确定度是组合的标准误差（ $U_c$ ），或这组成误差（随机的和系统的）的标准差，在测量过程中

乘以一个涵盖常数（K），这是对一期望置信度水准的正态曲线面积的表示；记住，应用了对测量过程呈正态分布的原则性假设。ISO/IEC《测量中不确定度指南》定义这涵盖常数是在正态分布下的 95% 面积来充分的呈现不确定度。通常认为  $K=2$ ：

$$U = KU_c$$

这组合标准差（ $U_c$ ）包括测量过程中所有重大变差的因素。在大多数情况下，按着本手册所实施的测量系统分析方法，可以被视为一种用来对测量不确定度的多种来源的定量工具。通常，可以将这最大误差要素定量为  $\sigma^2$  性能。其它较大的误差来源可以根据测量用途来应用。表达一不确定度时必须包括一个适当的范围，该范围能够识别所有的重要误差并且允许测量被重复进行。有些不确定度建立于长期测量系统误差，另一些是基于短期测量系统误差。然而，可以通过以下简单公式来计算：

$$U_c^2 = \sigma^2_{\text{性能}} + \sigma^2_{\text{其它}}$$

测量不确定度仅仅是用来估计在某测量时间下，测量可能的变化有多大，记住这一点很重要。应该考虑测量过程中所有重大的测量变差来源，加上校准、基准标准、方法、环境重大误差，以及其它在测量过程中没有预先考虑到的误差。在许多情况下，将使用 MSA 和 GRR 方法来计算那些重大的标准误差来估计不确定度。为了确保这不确定度估计值的持续准确性，对一测量过程相关的不确定度要适当周期性地再进行再评价。

---

<sup>26</sup>QS-9000，第3版，第4.11.1条

## 测量的不确定度和 MSA

不确定度和 MSA（测量系统分析）的主要区别是：MSA 专注于理解某测量过程，确定这过程中的误差大小，并评估这测量系统是否适用于产品和过程的控制。MSA 提升理解和改进（减少变差）。不确定度是测量值的范围，通过一个置信度区间的定义，与测量结果相关并预期包括测量的真值。

## 测量的可追溯性

可追溯性是通过一个不间断且描述了不确定度的比较链，对一测量特性或一标准数值与规定参考值的比较，这所指的通常是国家或国际标准；所以理解该链中每个环节的测量不确定度是很有必要的。正如在测量过程和追溯性链所介绍的，这包括短期的和长期的测量变差来源；因此测量系统的测量不确定度可被评价，以确保考虑了可追溯性的所有影响；接下来，这也能减少与测量相关的其它问题。

## ISO 表述测量不确定度的指南

*测量不确定度表达指南*（GUM）是如何评价并表达一测量的不确定度的指南。由于它对如何对测量不确定度的来源进行分类和组合，为使用者提供了理论上的解释和整套的指南；这应该被认为是一个高水平的参考文件，而不只是一个“如何使用”的手册。它的确给使用者提供了一些更进一步议题的指南，例如：变差来源的统计独立性、敏感度分析、自由度等，当评价更复杂、多参数测量系统时，这些议题是很关键的。

## 第一章— 第 G 节 测量问题分析

---

### 引言

理解测量变差以及它对整个变差的贡献是需要一个问题解决的基本步骤。当测量系统中的变差超过了其它任何变量，在对该系统进行其它分析研究之前，分析和解决这些问题便成有必要的工作。在有些情况下，造成测量系统的变差被忽视或忽略。这将会造成时间和资源的浪费，因为注意力被集中在过程本身，而事实上所呈现的变差是由测量装置所造成的。

本节，将对问题解决的基本步骤要进行探讨，并且说明它们的关系，以理解在一测量系统中存在的问题。每个公司可能都会使用这种顾客已经核准的问题解决过程。

如果测量系统是使用本手册介绍的方法所开发的，那么许多起始步骤将已经存在。例如早已经存在的因果图可能会对某测量过程提出有价值的经验学习。在任何正式的问题解决开始前，要能收集并评价这些资料。

## 第一步 识别问题

与研究其它任何过程一样，当研究测量系统时，清晰地定义问题或事件是很重要的。对于测量问题。可以从准确度、变差、稳定性等状态来考虑。最重要是试图从过程变差中，将测量变差及其影响个别分离出来（也就是说要研究过程，而不是研究测量装置）。问题的表述需要是一个适当可操作的定义，这样任何人都能理解解决这一问题。

## 第二步 识别小组

在这种情况下，问题解决小组将取决于测量系统和问题的复杂程度。一个简单的测量系统也许只需要几个人。但是，如果系统和问题很复杂，小组成员数也会增加（最多的小组人数要被限制在十个人以内）。小组成员以及他们代表的职能需要在这问题解决单上予以识别。

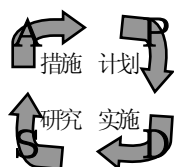
## 第三步 测量系统和过程的流程图

小组将要评审所有过去的测量系统和过程流程图。这样，可以促进对测量和与其有关的过程的已知和未知信息的讨论。过程流程图的分析可能会识别出需要加入其它人员到这分析小组。

## 第四步 因果图

小组将要评审在测量系统中所有过去的因果图。在某些情况下，这种评审可能会产生解决或部分解决问题的方法。这种评审还会促进对已经和未知的信息的讨论。小组将使用这些内容的知识初步的识别出对问题影响最大的那些变差。可能会执行额外的研究以证实这些决定。

## 第五步 计划—实施—研究—措施（PDSA）<sup>27</sup>



计划—实施—研究—措施是一个科学的研究形式。计划、试验、收集数据、建立稳定性、提出假设并证明，如此循环直到得到一个适宜的解决方法。

## 第六步 可能的解决方法及纠正的证明

步骤和解决方法被文件化，以记录这决定。进行初始研究以确认解决方法。可以使用一些试验设计的方法来进行解决方法的确认。另外，随时间的发展，还可能要进行额外的研究，包括环境和材料变差等。

## 第七步 将变更制度化

将最终解决方法记录在报告中，然后，由适当的部门或功能对过程进行适当改变，从而防止问题再发生。这样可能需要对程序、标准及培训材料作变更。这是整个问题解决过程中最重要的步骤。许多事件和问题在其它的时间会重复发生。

<sup>27</sup> 作者 W.爱德华·戴明，《产业、政府、教育的新经济》，MIT 印刷，1994，2000

## 第二章

### 用于评估测量系统的基本概念

## 第二章— 第 A 节 背景

### 引言

需要对两个重要领域进行评估：

1) 验证在适当的特性位置上测量了正确的变量。如适用，验证夹具和固定装置，同时识别与测量相互依赖的任何关键环境因素。如果一直在测量错误的变量，那么无论该测量系统多么准确或多么精确，都是在浪费资源，无任何益处。

2) 确定测量系统需要具备哪些可被接受的统计特征。为了做成这确定，知道将使用什么样的数据是很重要的，因为如果不了解这一点，就不能确定恰当的统计特性。统计特性确定之后，必须对测量系统进行评估，以确定实际上该测量系统是否具有这些特性。

### 第 1、2 阶段

**了解测量过程，以及该过程是否满足要求？**

第 1 阶段 试验是验证测量系统是否按照其设计规范要求，在适当的特性位置上测量正确的变量；（若适用，要验证夹具和固定装置）此外验证是否存在任何与测量相互依赖的重要环境问题。第 1 阶段可以使用一统计的设计实验来评估操作环境对测量系统参数的影响（例如偏倚、线性、重复性和再现性）。第 1 阶段实验结果可能表明操作环境不会对整个测量系统变差产生重大影响。另外，与重复性和再现性要素相比较，测量装置的偏倚和线性的影响应该较小。

从第 1 阶段试验中所得到的知识应该被用为开发测量系统维护计划的输入，同时应该被用在第 2 阶段试验类型的输入。环境问题可能驱使位置的改进，或对测量装置的环境控制。例如，如果整个测量系统变差中，重复性和再现性的影响很大，那么在第 2 阶段试验中可能要周期性简单的进行这两因素的统计实验。

**随时间的推移，测量系统是否能持续满足要求？**

第 2 阶段 试验是对变差的主要原因提供持续的监控，从而说明测量系统是持续可信的（以及所产生的数据）和/或随着时间推移，测量系统是否出现变坏的信号。



## 第二章— 第 B 节 选择/开发试验程序

“如果理解和遵守技术的极限，则任何技术都可能是有用的。”<sup>28</sup>

有许多适当的程序可用于评估测量系统。选择使用哪些程序取决于许多因素，其中许多因素需要其于被评估的每个测量系统的个案来逐一确定。在某些情况下，为确定一个程序对一特定的测量系统是否合适，可能需要预先的试验。这种预先试验要上一章节讨论中的第一阶段试验的一个完整的部分。

当选择或开发一个评估程序时，一般要考虑的问题包括：

- 试验中是否应使用诸如那些可追溯至 NIST 的标准？如果是，什么等级的标准是合适的？标准经常是评估一个测量系统的准确度所必要的。如果不使用标准，该测量系统的变异性仍有可能被评估，但不大可能按合理的可信度去评估该系统的准确度。缺乏这样的可信度可能是一个问题，例如，试图解决一个生产者的测量系统和一个顾客测量系统之间明显的差别；
- 对于第 2 阶段持续的试验，应考虑使用盲测。盲测法是指在实际测量环境下，由一事先不知正在对该测量系统进行评估的操作者所获得的测量结果；通过适当的管理，根据得到的试验结果通常不受众所周知的霍桑效应所干扰；<sup>29</sup>
- 试验成本；
- 试验所需要的时间；
- 任何没有被普遍接受的术语应被可操作的定义。这些术语如准确度、精确度、重复性和再现性等；
- 是否由这个测量系统取得的测量结果要与另外一个测量系统得到的测量结果对比？如果对比，应考虑使用依赖于使用诸如前面第 1 阶段中所讨论的标准。如果不使用标准，仍有可能确定两个测量系统是否可以同时正常工作。然而，如果两个系统一起工作不正常，那么不用标准，就不可能确定哪个系统需要改进；
- 第 2 阶段试验应每隔多久进行一次？这个问题可能基于个别测量系统的统计特性及其对设施后续结果来决定；以及以该设施进行制造过程，但事实由于一个测量系统没有正常工作而未受监控的制造过程顾客来决定。

除了这些一般性的问题外，正被试验的特殊测量系统的其它特定问题也可能是重要的。发现特殊测量系统的具体问题是第 1 阶段试验的两个重要目的之一。

---

<sup>28</sup> 爱德华 戴明，*评价的逻辑*，评价研究手册，第一卷，编者：Elmer L. Struening 和 Marcia Guttentag.

<sup>29</sup> “霍桑效应”是指 1924 年 11 月到 1932 年 8 月间，在西部电气公司的霍桑工厂 (the Hawthorne Works of Western Electric) 完成的一系列工业试验的结果。在试验中，研究人员系统地变更了五个装配工的工作条件，并监视结果。由于条件的改善，产量上升，然而，当工作条件下降时，产量继续增长，这仅仅是因为这些工人是这项研究工作的一部分，由此而使他们产生了更积极的工作态度的结果，而不是改变了工作条件的结果。更详细的信息，参见由 Richard Gillespie 编写的《〈霍桑实验史〉》，剑桥大学出版社，纽约 1991。

## 第二章— 第 C 节 测量系统研究的准备

如任何研究或分析一样，进行测量系统研究之前要进行充分的策划和准备。进行测量系统研究的准备一般包括：

- 1) 应该策划所使用的方法。例如，如果在校准或使用仪器时有评价人的影响，使用工程判断、目视观察或量具研究来确定。有些测量系统的再现性影响可以忽略，例如只需按一下按钮，测量结果就能打印出来的测量系统；
- 2) 应该事先确定评价人的人数、抽样零件的数量，及重复读数的数量等。选择以上内容时要考虑的因素如下：
  - (a) 尺寸的关键性 — 关键尺寸需要更多的零件和/或测量次数。原因是量具研究估计用的置信度需求；
  - (b) 零件的形态 — 大型或重型零件可能意味着样本较少但测量次数多。
  - (c) 客户设备
- 3) 由于目的是评价全部的整个测量系统，评价人应该从那些正常操作该食品的人员中选择
- 4) 样品的选择对适当的分析很关键的，其完全的取决于 MSA 研究的设计、测量系统的目的及代表该生产过程的零件样本的可获得性。

对于产品控制情况下，当测量结果与决定准则是确定，“符合或不符合某特性的规范”（如 100%检验或抽样），样品（或标准）必须被选择，但不需要包括整个过程范围。测量系统的评估是特性公差为基础（如相对公差的%GRR）。

在过程控制情况下，测量结果与决定准则是确定“过程稳定性，方向和符合自然过程变差（如 SPC、过程监视、能力及过程改进），整个作业过程范围的样本可获得性变得非常重要。当评估一测量系统对过程控制的适用性时（如对过程变差的%GRR），推荐采用过程变差的独立估算法（过程能力研究）。

当不能使用过程变差的独立估算法时、或为了确定过程方向和测量系统对过程控制的持续适宜性时，必须从过程中选择样件，并且该样件能代表整个生产作业范围。使用 MSA 研究中所选择的抽样零件的变差（PV）来计算这研究的总变差（TV）。TV 指数（如相对 TV 的%GRR）是对过程方向以及对过程控制的测量系统的持续适宜性的一个指标值。若样品不能代表生产过程，评定中必须忽略 TV。忽略 TV 不会影响到使用公差（产品控制）的评定或一个过程变差的独立估算法（过程控制）。

零件抽样的选择可以是在许多天中每天抽一个样品。再者，这是必要的，因为零件代表生产过程中生产变差的范围，这些零件将用来作为分析。由于每一个零件将被测量若干次，必须对每一个零件编号以便于识别。

- 5) 仪器应该有足够的分辨力，特性直接被读出的值至少是预期过程变差的十分之一。例如，

如果特性的变差为 0.001，仪器应能读取 0.0001 的变化；

6) 确保测量方法（如评价人和仪器）是测量特性的尺寸，并遵循已定义的测量程序。

进行研究的方式十分重要。本手册介绍的所有分析方法均假设单个数值之间具有统计上的独立性<sup>30</sup>。为了减少误导结果的可能性，需采取以下步骤：

(1) 测量应按照随机顺序<sup>31</sup>进行，从而确保可能发生的漂移或变化将随机分布在整个研究中。评价人应该不能察觉正在检查的零件编号，以避免任何可能因了解所成的偏倚。但是，主导研究的人应知道正在检查哪一零件，并相应记下数据，即评价人 A，零件 1，第一次测量；评价人 B，零件 4，第二次测量等。

(2) 对设备读值，应该按测量仪器解析度的实际极限纪录这测量值。机械式装置必须读到并记录到最小的刻度单位。对于电子读数，测量计划中必须规定共同原则以记录准确的有意义显示数值。对于类比式装置，应该记录最小刻度或灵敏度与分辨率极限值为 0.0001 时，则测量结果应该被记录到 0.00005 时。

(3) 这研究工作应该由了解实施一个可靠研究的重要性的人员负责管理和监督。

当开发第 1 阶段或第 2 阶段试验计划时，考虑以下因素：

- 评价者对测量过程有哪些影响？若有可能，研究中应该包括正常使用该测量装置的评价者。

每个评价人都应该使用程序— 包括所有步骤---通常他们用来获取读值。评价人所使用方法之间的不同将影响测量系统的再现性。

- 评价者对测量设备的校准是否可能是一重大的变差因素？如果是这样，评价者在每个读值分组之前应重新校准设备。

- 需要多少个取样零件和重复几次读值？所要求的零件的数量将取决于被测特性的重要性以及测量系统变差估计中所要求的置信度水准。

当使用本手册推荐的方法进行研究时，尽管评价者人数、测量次数以及零件数量会有所不同，在第 1 阶段和第 2 阶段试验计划之间、或为同一个测量系统的连续第 2 阶段试验之间，通常保留使用相同的评价者人数、试验次数和零件数量。在试验计划之间及连续的试验之间保持一致，将提高在不同的试验结果之间的可比较性。

<sup>30</sup> 数之间没有相互关系。

<sup>31</sup> 见第三章，第二节，“随机化及统计独立性”

## 第三章— 第 D 节

### 结果分析

应该对结果进行评价，以确定该测量装置就其预期的应用是否可接受。一个测量系统在任何附加的分析生效之前应该是稳定的。

### 装备或夹具

接受性准则---量具---装备和夹具误差

设计不合理的夹具或安装不当的量具会增加测量的误差。而这种情况经常在测量过程不稳定或不可控制的环境下发生。这种误差可能是由于量具偏差过大或由于量具重复性低及准确性差的 GRR 值引起的。

通常情况下，如存在明显的测量问题时，首先应该做的是查阅装配和安装指南以确保量具是否合理安装，（注：某些问题可能指南中未说明），如夹具/探针应合理定们及适当负重。同时对自动化测量系统而方，首先必须检查整个方案是否符合要求或预期草案。

如其他区域发生故障必须重新安装或修理量具及夹具，然后重新运行测量评估系统。

### 位置误差

接受性准则 — 位置误差

位置误差通常被定义为偏倚和线性的分析。

通常，如果一测量系统的偏倚或线性误差明显的异于零、或是超出量具校准程序确立的最大允许误差，那么它是不可接受的。在这种情况下，应对测量系统重新进行校准或进行一个补偿修正来减少这误差。

### 宽度误差

接受性准则 — 宽度误差

一测量系统变差是否为满意的准则，取决于测量系统变差对制造过程误差或零件公差所占的百分比。对特定的测量系统的最终接受准则取决于测量系统环境和目的，而且应该取得顾客的同意。

对于以分析过程为目的的测量系统，通常单凭经验来确定测量系统的可接受性的规则如下：

- 误差低于 10% — 通常认为测量系统是可接受的。
- 误差在 10%到 30%之间 — 根据应用的重要性、测量装置的成本、维修费用等方面的考虑，可能是可接受的。
- 超过 30%的误差 — 认为是不可接受的 — 应该尽各种力量来改进测量系统。

此外，由测量系统对过程进行划分的区别分类数（ndc）<sup>32</sup>要能大于或等于 5。

---

<sup>32</sup> 第三章，第二节“数据结果的分析”

一测量系统的最终可接受性应该不仅仅取决于一些简单的指数，应该同时使用随时间变化的图表来分析测量系统的长期性能。

## 第三章

### 对简单测量系统的推荐实施方法

#### 第二章 第 A 节

#### 试验程序范例

#### 引言

本章将介绍一些特定的试验程序范例。这些程序简单实用并可容易的使用于一生产环境。如前面所讨论的那样，试验程序应该被用来理解一测量系统并量化其变差，并取决于那些可能影响测量系统的变差来源。在许多情况下，主要的变差来源是由于仪器（量具/设备），人（评价人）和方法（测量程序）造成的。本章介绍的试验程序是充分的的对这种类型的测量系统分析。

程序适用于当：

- √ 只研究两个因素，或者称为测量条件（如评价人和零件）加上所研究的测量系统重复性。
- √ 每个零件骨的变异性的影响可以忽略
- √ 不存在统计上的评价人和零件之间的交互作用
- √ 在研究中零件的尺寸不发生变化

可以进行试验统计设计和/或用相关专业知识来判断这些程序是否适用于任何特定的测量系统。

## 第四章— 第 B 节

### 计量型测量系统研究----指南

#### 引言

本节包括了在第一章第 E 节所描述的测量系统技术实施指南。建议在适当的应用这些指南之前，彻底复习一遍第 E 节的内容。

#### 确定稳定性的指南

##### 进行研究

- 1) 取一样件并建立其可追溯到相关标准的参考值。如果该样品无法取得，选择一个落在产品测量中间的生产零件<sup>33</sup>，指定其为基准样件以进行稳定性分析。追踪测量系统稳定时，不要求该已知的参考值。

这也许会希望拥有位于预期测量结果的下限、上限和中间位置的基准件。推荐对每种基准件单独的进行测量和画控制图。

- 2) 以一定的周期基础（天、周）测量基准件 3~5 次，抽样数量和频率应该取决于对测量系统的认识。可能考虑的因素包括重新校准或维修的频率如何、测量系统的频率，以及操作条件的重要性等。应在不同的时间下取得多次读值，以代表测量系统的实际使用情况，这将考虑了在一天之中因为热机、周遭或其它因素可能发生的变化。

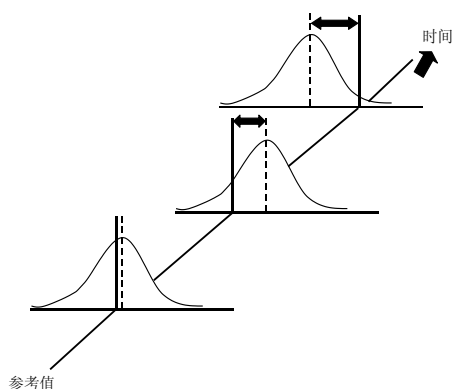
- 3) 将数据按时间顺序画在  $\bar{X}$ &R 或  $\bar{X}$ &S 控制图上

##### 结果分析-图示法

- 4) 建立控制限，使用控制图分析法来评价是否有不受控或不稳定情况。

##### 结果分析-数值法

除了正常控制图分析法外，没有用于稳定性分析特别数值分析或指数<sup>34</sup>。



<sup>33</sup> 应该注意，由于使用、材料和搬运的原因可能造成生产基准零件过度磨损。这将要求对该生产零件进行某些改变，如电镀，以延长基准零件的寿命。

<sup>34</sup> 见 SPC 参考手册。

如果测量过程是稳定的，数据可以用于确定测量系统的偏倚。

另外，测量的标准差可以用作测量系统重复性的近似值。重复性可以与过程的标准偏差相比较，以确定测量系统的重复性是否适用于应用。

使用实验设计或其它问题解决的分析技术可能是必要的，以确定测量系统缺乏稳定性的主要原因。

### 范例-稳定性

为了确定一个新的测量仪器稳定性是否为可以接受，过程小组选取了在生产过程输出范围中接近中间值的一个零件。这个零件被送到测量实验室，经测量其参考值确定为 6.01。小组每班测量这个零件 5 次，共测量 4 周（20 个子组）。收集所有数据以后，X&R 图就可以做出来了（见图 9）

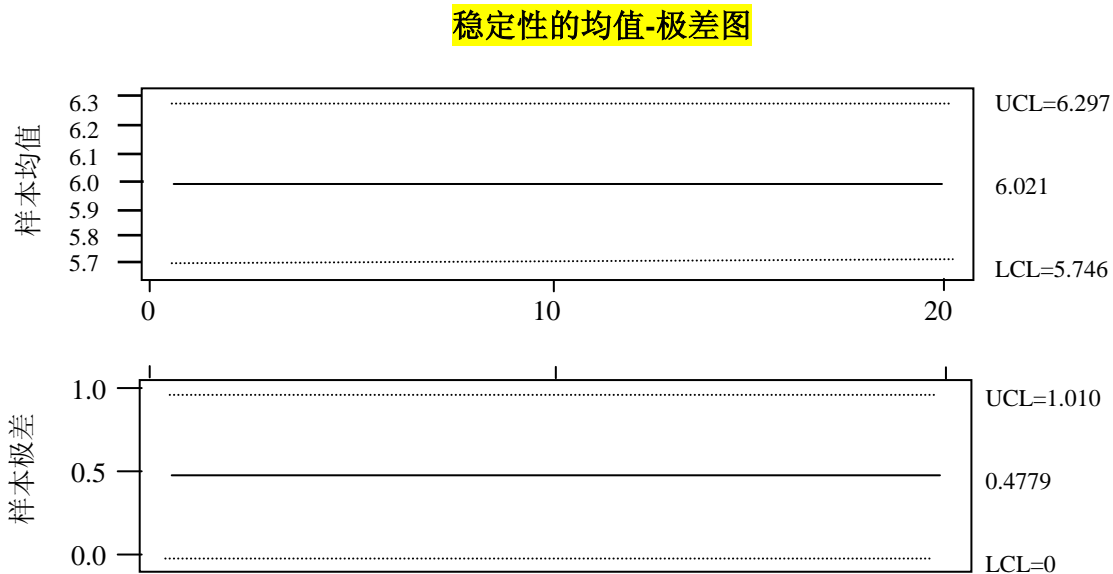


图 9：用于稳定性分析的控制图

控制图分析显示，测量过程是稳定的，因为没有出现明显可见的特殊原因结果发生。

### 确定偏倚指南<sup>35</sup> —独立样件法

#### 进行研究

1) 获取一个样件，并建立与其可追溯性到相关标准的参考值。如果得不到参考值，选择一个落在生产测量范围中间的生产件，指定其为偏倚分析的基准件。在工具室测量这个零件  $n \geq 10$  次，并计算这  $n$  个读数的均值。把均值作为“参考值”。

<sup>35</sup> 见第二章，第五节，用于操作的定义和潜在的原因讨论

也许会希望拥有位于期望测量结果的下限、中间及上限位置的基准条件,如果可做到这样,可以使用线性研究来分析这些数据。

2) 让一个评价人,以通常方法测量样本 10 次以上,

结果分析---图示法

3) 画出这些数据相对于参考值的直方图。使用专业知识评审这直方图,从而确定是否存在任何特殊原因或异常点。如果不存在,继续分析,当  $n < 30$  时,对任何的解释或分析时,要能特别注意。

结果分析-数值法

4) 计算  $n$  个读数的均值。

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

5) 计算重复性标准差(参考下面的量具研究,极差法):

$$\sigma_{\text{重复性}} = \frac{\max(x_i) - \min(x_i)}{d_2^*}$$

这里  $d_2^*$  可以从附录 C 中查到,  $g=1, m=n$

如果可以获得 GRR 研究(且有效),重复性标准偏差计算应该取决于这研究结果。

6) 确定偏倚的  $t$  统计量<sup>36</sup>:

$$\text{偏倚} = \text{观测到的测量平均值} - \text{参考值}^{37}$$

$$\sigma_b = \sigma_r / \sqrt{N}$$

$$t = \frac{\text{偏倚}}{\sigma_b}$$

<sup>36</sup> 这种方法使用的平均极差接近于测量过程的标准偏差。因为极差点的使用的有效性随样本容量的增加而增大,如果总的样本容量超过 20,建议是将样本分成多个子组并使用控制图法,或使用传统的一个样本  $t$  试验得到标准偏差的均方差 RMS 计算。

<sup>37</sup> 偏倚的不确定度由  $\sigma_b$  给出。



7) 如果 0 落在偏倚值附近的 1-a 置信度界线内, 则偏倚在 a 水准上是可接受的。

$$\text{偏倚} - \left[ \frac{d_2 \sigma_b}{d_2^*} (t_{v, 1-a/2}) \right] \leq 0 \leq \text{偏倚} + \left[ \frac{d_2 \sigma_b}{d_2^*} (t_{v, 1-a/2}) \right]$$

公式中, v 可以在附录 C 中查到  $t_{1-a/2}$  可以利用标准 t 分布表中查到。

所使用的 a 水准取决于敏感度水准, 敏感度水准对评价/控制一过程是必要的, 并且与产品/过程的损失函数 (敏感度曲线) 有关。如果 a 水平不是用预设值 0.05 (95% 置信度) 则必须得到顾客的同意。

#### 偏倚—范例

一个制造工程师评价了一个用来监视生产过程的新的测量系统。测量设备的一项分析表明该测量系统应该没有线性误差的考量, 所以该工程师只需对测量系统的偏倚进行评价。他基于一份已经文件化的过程变差描述, 在这测量系统操作范围内选取了一个零件; 通过对该零件进行了全尺寸测量来确定它的参考值, 然后由主要操作者测量该零件 15 次。

测量次数	参考值=6.0	偏倚
1	5.8	-0.2
2	5.7	-0.3
3	5.9	-0.1
4	5.9	-0.1
5	6.0	0.0
6	6.1	0.1
7	6.0	0.0
8	6.1	0.1
9	6.4	0.4
10	6.3	0.3
11	6.0	0.0
12	6.1	0.1
13	6.2	0.2
14	5.6	-0.4
15	6.0	0.0

表 2: 偏倚研究数据

通过使用和统计软件, 检验员得到了直方图和数值分析结果(见图 10 和表 3)。

测量值

图 10: 偏倚研究- 偏倚研究直方图

由于 0 落在偏倚置信区间 (-0.1215,0.1349)，工程师可以假设测量偏倚是可以接受的，即在实际使用时,将不会带来额外的变差来源.

	n (m)	均值 $\bar{X}$	标准偏差 $\sigma_r$	均值的标准偏差 $\sigma_b$
测量值	15	6.0067	.22514	.05813

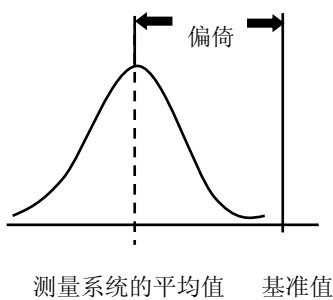
	基准值 = 6.00, a = .05 g = 1, $d_2^* = 3.35$				
	t 统计量	df	显著 t 值 (2 尾)	偏倚	95% 偏倚置信区间
					低值      高值
测量值	.1153	10 .8	2.206	.0067	-1.1185      .1319

表 3: 偏倚研究-偏倚研究的分析

确定偏倚的指南-控制图  
进行研究

如果均值-极差图或均值标准差图用于测量稳定性，其数据也可以用来评价偏倚。在评价偏倚之前，控制图分析应该表明这测量系统处于稳定状态。

1) 获取一个样件并建立于可追溯到一相关标准的参考值。如果不能取得参考值，选择一个落在产品测量值范围中间的生产零件,将它指定为偏倚分析的基准样件。在工具室里测量该零件  $n \geq 10$  次，并计算这 n 个数据的均值。把均值作为“参考值”。



结果分析-图示法

2) 画出这些数据相对于参考值的直方图。评审直方图，以专业知识确定是否存在特殊原因变差或出现异常。如果没有，继续这项分析。

结果分析-数值法

3) 从控制图得到平均值  $\bar{X}$

4) 用平均值  $\bar{X}$  减去参考值,计算得到偏倚

$$\text{偏倚} = \bar{X} - \text{参考值}$$

5) 用极差的平均值计算重复性标准差

$$\sigma_{\text{重复性}} = \frac{\overline{R}}{d_2^*}$$

公式中,  $d_2^*$  取决于子组大小(m)和在控制图中子组数量(g)。(见附录 C)

6) 确定对偏倚的统计 t 值<sup>38</sup>:

$$\sigma_b = \sigma_r / \sqrt{mg}$$

$$t = \frac{\text{偏倚}}{\sigma_b}$$

7) 如果 0 落在偏倚值附近的 1-a 置信区间内, 则偏倚在这 a 水平内可接受。

$$\text{偏倚} - \left[ \frac{d_2 \sigma_b}{d_2^*} (t_{v, 1-a/2}) \right] \leq 0 \leq \text{偏倚} + \left[ \frac{d_2 \sigma_b}{d_2^*} (t_{v, 1-a/2}) \right]$$

这里  $d_2$ ,  $d_2^*$  和 v 可以在附录 C 中查到,  $t_{v, 1-a/2}$  在标准 t 表中查到。

所使用的 a 水准取决于敏感度水准, 敏感度水准对评价/控制一过程是必要的, 并且与产品/过程的损失函数(敏感度曲线)有关。如果 a 值置信度水准不是使用预设值 0.05(95%置信度), 则应该得到顾客的同意。

### 范例-偏倚

参考图 9, 对一个参考值为 6.01 的零件进行了稳定性研究, 所有样本(20 个子组)的整体平均值为 6.021。因而计算偏倚值为 0.011。

使用散布图和统计软件, 生成的数据分析见表 4。

由于 0 落在偏倚的置信度区间内 (-0.0299, 0.0519), 该过程小组可以假设这测量系统的偏倚是可以接受的, 即在实际使用中不会带来额外的变差来源。

	n (m)	均值 $\bar{X}$	标准偏差 $\sigma_r$	均值的标准偏差 $\sigma_b$
测量值	100	6.021	.2048	.02048

	参考值= 6.01, a= .05, m=5, g=20, $d_2^* = 2.334$ , $d_2 = 2.326$					
	t 统计量	df	显著 t 值 (2 尾)	偏倚	95% 偏倚置信区间	
					低值	高值
测量值	.5371	72.7	1.996	.011	-.0299	.0519

表 4: 偏倚研究-对偏倚的稳定性研究的分析

<sup>38</sup> 偏倚的不确定度由  $\sigma_b$  给出。

## 偏倚研究的分析：

如果偏倚在统计上不等于 0，检查是否存在以下原因：

- 基准件或参考值有误，检查确定标准件的程序
- 仪器磨损。这问题会在稳定性分析中呈现出来，建议进行维护或重新修正计划
- 仪器产生错误的尺寸
- 仪器所测量特性有误
- 仪器没有经过适当的校准，对校准程序进行评审
- 评价者使用仪器的方法不正确，对测量指导书进行评审。
- 仪器纠正的指令错误

如果测量系统偏倚不等于零.若有可能,应该采用硬件修正法、软件修正法或同时使用这两种方法来对量具进行重新校准以达到 0 偏倚,如果偏倚不能调整到 0,通过变更程序(如:对每个读值根据偏倚进行修正)还可以继续使用该系统。由于存在评价人误差这一高度风险,因此这种方法只能在取得顾客同意后方能使用。

---

## 确定线性的指南<sup>39</sup>

进行研究

可使用以下指南进行线性评价：

- 1) 由于存在过程变差,选择  $g \geq 5$  个零件,使这测量值涵盖量具的整个工作量程。
  - 2) 对每个零件进行全尺寸检验测量,以确定其参考值,并确涵盖了这量具的工作量程。
  - 3) 让经常使用该量具的操作者测量每个零件  $m \geq 10$  次。
- √ 随机地选择零件,从而减少评价人对测量中偏倚的“记忆”。

---

<sup>39</sup> 参见第一章第 E 节,对潜在原因的可操作性定义及讨论

结果分析- 作图法

4) 计算零件每次测量的偏倚,以及每个零件的偏倚均值。

$$\text{偏倚}_{I, j} = x_{I, j} - (\text{参考值})_I$$

$$\overline{\text{偏倚}}_I = \frac{\sum_{j=1}^m \text{偏倚}_{I, j}}{m}$$

5) 在线性图上画出相对于参考值的每个偏倚和偏倚平均值 (见图 11)。

6) 用下面等式计算和画出最适合的线和该线的置信度区间。

对于最适合的线, 用公式:  $\overline{y}_i = ax_i + b$

$x_i =$  基准值

$y_i =$  偏倚平均值

这里  $x_i$  是基准值,  $y_i$  是偏倚均值, 并且

$$\text{公式: } a = \frac{\sum xy - \left\{ \frac{1}{gm} \sum x \sum y \right\}}{\sum x^2 - \frac{1}{gm} (\sum x)^2}$$

$$b = \overline{y} - a\overline{x} = \text{截距}$$

对于给定的  $x_0$ ,  $a$  水平置信度区间<sup>40</sup>为:

$$s = \sqrt{\frac{\sum y_i^2 - b \sum y_i - a \sum x_i y_i}{gm - 2}}$$

$$\text{下限: } b + ax_0 - \left[ t_{gm-2, 1-a/2} \left\{ \frac{1}{gm} + \frac{(x_0 - \overline{x})^2}{\sum (x_i - \overline{x})^2} \right\}^{1/2} s \right]$$

$$\text{上限: } b + ax_0 + \left[ t_{gm-2, 1-a/2} \left\{ \frac{1}{gm} + \frac{(x_0 - \overline{x})^2}{\sum (x_i - \overline{x})^2} \right\}^{1/2} s \right]$$

<sup>40</sup> 参见第三章第 B 节, 确定偏倚的指南中  $a$  程度选择中的注释。

7) 画出“偏倚=0”线，并对图进行评审,以观察是否存在特殊原因,以及线性是否可接受。(见例图 11)

如果“偏倚=0”的整个直线都位于置信度区间以内,则称该测量系统的线性是可接受的。

### 结果分析-数值法

8) 如果图示法分析表示该测量系统线性可接受，则下面的假设就成立：

$$H_0: a=0 \quad \text{斜率}=0$$

如果下式成立,则不能被否定

$$|t| = \frac{\frac{|a|}{s}}{\left[ \frac{1}{\sum (x_i - \bar{x})^2} \right]} \leq t_{gm-2, 1-a/2}$$

如果以上的假设是成立的,则测量系统对所有的参考值具有相同的偏倚.对于可接受的线性,偏倚必须为 0。

$$H_0: b=0 \quad \text{截距 (偏倚)} =0$$

如果正式成立,则不能被否定

$$|t| = \frac{|b|}{\left[ \sqrt{\frac{1}{gm} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} \right]}$$

### 范例- 线性

某工厂检验员对某过程引进了一套新测量系统。作为 PPAP<sup>41</sup> 的一部分需要对测量系统的线性进行评价。根据已文件化的过程变差描述,在测量系统的全部工作量程范围内选取了五个零件.通过对每个零件进行全尺寸检验测量以确定其参考值。然后由主要操作者对每个零件测量 12 次。在分析中,这些零件是随机抽取的。

<sup>41</sup> 生产件批准程序手册，第三版，2000。

	零件 参考值	1	2	3	4	5
		2.00	4.00	6.00	8.00	10.00
试 验	1	2.70	5.10	5.80	7.60	9.10
	2	2.50	3.90	5.70	7.70	9.30
	3	2.40	4.20	5.90	7.80	9.50
	4	2.50	5.00	5.90	7.70	9.30
	5	2.70	3.80	6.00	7.80	9.40
	6	2.30	3.90	6.10	7.80	9.50
	7	2.50	3.90	6.00	7.80	9.50
	8	2.50	3.90	6.10	7.70	9.50
	9	2.40	3.90	6.40	7.80	9.60
	10	2.40	4.00	6.30	7.50	9.20
	11	2.60	4.10	6.00	7.60	9.30
	12	2.40	3.80	6.10	7.70	9.40

表 5: 线性研究数据

使用散布图和统计软件，检查员画出了线性图（图表 11）。

	零件 参考值	1	2	3	4	5
		2.00	4.00	6.00	8.00	10.00
偏 倚	1	0.70	1.1	-0.2	-0.4	-0.9
	2	0.50	-0.1	-0.3	-0.3	-0.7
	3	0.40	0.2	-0.1	-0.2	-0.5
	4	0.50	1	-0.1	-0.3	-0.7
	5	0.70	-0.2	0.0	-0.2	-0.6
	6	0.30	-0.1	0.1	-0.2	-0.5
	7	0.50	-0.1	0.0	-0.2	-0.5
	8	0.50	-0.1	0.1	-0.3	-0.5
	9	0.40	-0.1	0.4	-0.2	-0.4
	10	0.40	0.0	0.3	-0.5	-0.8
	11	0.60	0.1	0.0	-0.4	-0.7
	12	0.40	-0.2	0.1	-0.3	-0.6
偏倚平均值		0.491667	0.125	0.025	-0.29167	-0.61667

表 6: 线性研究- 中途的结果

线性例子

$$Y=0.736667 - 0.131667X$$

$$R-Sq = 71.4\%$$

## 基准值

图 11: 线性研究-作图分析

图示法分析指出特殊原因可能会影响测量系统。参考值为 4 数据呈现两种统计模式。即使不考虑参考值为 4 的数据，图示分析明显地表示该测量系统存有一线性问题。 $R^2$  值指出线性模型可能不是这些数据的适当模型<sup>42</sup>。即使该线性模式是可接受的，这“偏倚=0”的直线与置信度区间相交,而不是包含在置信度区间内。

此时，这检验员要开始分析并解决测量系统存在的问题，因为数值分析不能提供任何其进一步的方向。然而，为确保不能半途而废,检验员对斜率和截距计算统计的 t 值.

$$t_a = -12.043$$

$$t_b = -10.158$$

取预设  $\alpha = 0.05$ ，由 t 分布表,取  $(gm - 2) = 58$  个自由度,并取概率为 0.975,该督导者得出关键值:

$$t_{58,.975} = 2.00172$$

由于  $|t_a| > t_{58,.975}$  ,由图示法分析得到的结论经过了数值法分析的证实----本测量系统存在线性的问题。

在这种情况下，由于测量系统存在一线性问题，不用再去分析  $t_b$  与  $t_{58,.975}$  的关系如何。引起线性问题可能的原因可以在第一章第 E 节“位置变差”。

如果测量系统存在线性问题，需要通过硬件修正、软件修正、两项同时进行来再校准以达到 0 偏倚。

如果不能将整个测量系统范围内偏倚调整到零偏倚，它还是可能被用为对产品/过程的控制，只要这测量系统依旧稳定，可以不加以分析。

由于这有一个评价者的误差风险较高，该测量应该仅在得到顾客同意下被使用。

<sup>42</sup> 见标准统计课程关于使用线性模型来描述两个变量的关系的适当分析。



## 确定重复性和再现性的指南<sup>43</sup>

可以使用不同的方法进行计量型量具的研究。本节将详细讨论三种可接受的方法。它们是：

- 极差法
- 均值极差法
- 方差分析法

除了极差法，其他方法所用的研究数据设计都很相似。如所呈现的，所有方法在它们分析时均忽视了零件内的变差（如在第四章，第 A 节所讨论的圆度、锥度直径、平面度等）。

但是，整个测量系统包括的不仅是量具本身和及其相关的偏倚、重复性等，还包括被测量零件之间的变差。如何处理零件内的变差，需要取决于对零件使用意图以及测量目的的合理解释。

最后，本节中所有技术都是以过程处于统计的稳定状态这一前提条件。

*尽管再现性通常被解释为评价人变差，但有些情况下该变差会由于其它原因造成。例如，对于一些线上没有评价人的测量系统，如果所有的零件由相同的设备来搬运、夹紧及测量，则再现性为零。在这范例中只有对重复性研究的需要。但是，如果使用多夹具，则存在多种夹具之间的再现性变差。*

### 极差法

极差法是一种经修正的计量型量具的研究方法，它能对测量变差提供一个快速的近似值。这种方法只能对测量系统提供变差的整体情况，不能将变差分解成重复性和再现性。它通常用来快速地检查以验证 GRR 是否有变化。

使用这个方法能够潜在的检测出测量系统为不可接受<sup>44</sup>的概率是：对于抽样次数是 5 的情况下，概率为 80%，对于抽样次数为 10 的情况下，概率为 90%。

用极差法进行研究时通常选用两个评价人与五个零件。在这种研究中，两个评价人测量每个零件一次。由评价人 A 测量的每个零件的极差与由评价人 B 测量的每个零件的极差是决然不同的。计算极差之和以及极差的平均值；总测量变差即为极差的平均值乘以  $1/d_2$ 。  $d_2^*$  在附录 C 中可以找到，  $m=2$ ，  $g=$  零件件数。

<sup>43</sup> 见第一章，第五节，可操作的定义和潜在原因的讨论。

<sup>44</sup> 如 %GRR > 30%。

零件	评价人 A	评价人 B	极差 (A, B)
1	0.85	0.80	0.05
2	0.75	0.70	0.05
3	1.00	0.95	0.05
4	0.45	0.55	0.10
5	0.50	0.60	0.10

$$\text{极差平均值(R)} = \frac{\sum R_i}{5} = \frac{0.35}{5} = 0.07$$

$$\text{GRR} = \left( \frac{\bar{R}}{d_2^*} \right) = \left( \frac{0.07}{1.19} \right) = 0.0588 \approx 0.06$$

(从前的研究可知:过程标准偏差=0.0777)

$$\% \text{GRR} = 100 * \left( \frac{\text{GRR}}{\text{过程标准偏差}} \right) = 75.7\%$$

表 7: 量具研究 (极差法)

为了确定测量变差占过程标准差的百分比, 通过将 GRR 乘以 100 除以过程标准差, 将其转换为百分比。在例子 (见表 7) 中, 这个特性的过程标准偏差是 0.0777, 因而:

$$\% \text{GRR} = 100 * (\text{GRR} / \text{过程标准偏差}) = 75.7\%$$

现在测量系统的 %GRR 已经确定, 应该进行结果的解释。在表 7 中, %GRR 确定为 75.7%, 结论是测量系统需要改进。

### 平均值和极差法

平均值和极差法 (X&R) 是一种可同时对测量系统提供重复性和再现性的估计值的研究方法。与极差法不同, 这方法允许将测量系统的变差分解成两个独立的部分----重复性和再现性, 但是不能确定它们两者的交互作用<sup>45</sup>。

进行研究

尽管评价人的人数、测量次数和零件数均可会不同, 但下面的讨论呈现进行研究的最佳情况。参见图 12 中的 GRR 数据表。详细的程序是:

<sup>45</sup> 方差分析法可以用来确定量具和评价人之间的交互作用, 如果存在这种情况。

- 1) 获取一个能代表过程变差实际或预期范围的样本，零件数  $n > 5$  个零件的样本<sup>46</sup>
- 2) 给评价人编号为 A, B, C 等。并将零件从 1 到  $n$  进行编号，但零件编号不要让评价人看到。参见第 2 章，第 C 节。
- 3) 对量具进行校准，如果这是正常测量系统程序中的一部分的话。让评价人 A 以随机顺序测量  $n$  个零件<sup>47</sup>，将测量结果输入第一行。
- 4) 让评价人 B 和 C 依次测量这些一样的  $n$  个零件，不要让他们知道别人的读值；然后将结果分别的记录在第 6 行和第 11 行。
- 5) 用不同的随机测量顺序重复以上循环。并将数据记录在到第 2, 7 和 12 行。注意将数据记录在适当的栏位中。如果首先被测量的是第 7 号零件，然后将数据记录在标有零件 7 的栏位中。如果需要进行三次测量，则重复以上循环，并将数据记录在第 3, 8 和 13 行中。
- 6) 当测量大型零件或不可能同时获得数个零件时，第 3 步到第 5 步将变更成以下顺序：
  - √ 让评价人 A 测量第一个零件并在第 1 行记录读数。让评价人 B 测量第一个零件并在第 6 行记录读数。让评价人 C 测量第一个零件并在第 11 行记录读数。
  - √ 让评价人 A 重复测量第一个零件并记录读数于第 2 行，让评价人 B 重复测量第一个零件并记录读数于第 7 行，让评价人 C 重复测量第一个零件并记录读数于第 12 行，如果试验需要进行 3 次，重复这个循环将数据记录在第 3, 8, 13 行。
- 7) 如果评价人处于不同的班次，可以使用一个替代方法。让评价人 A 测量所有的 10 个零件输入数据于第 1 行，然后评价人 A 以不同的顺序读数重新测量，并把记录结果记录在于第 2, 3 行，让评价人 B, C 同样做。

---

<sup>46</sup> 对于在结果中最低的置信水平，总的极差数量应该大于 15。尽管表格被设计为最多 10 个零件，这种方法不被序号所限制。如同任何统计技术，较大的样本容量，将表现出减少抽样变差和结果的风险。

<sup>47</sup> 见第三章，第二节，“随机化及统计独立性”。

量具重复性和再现性数据收集表

评价人/ 测量次数	零件										平均值
	1	2	3	4	5	6	7	8	9	10	
A 1	0.29	-0.56	1.34	0.47	-0.8	0.02	0.59	-0.31	2.26	-1.36	
2	0.41	-0.68	1.17	0.5	-0.92	-0.11	0.75	-0.20	1.99	-1.25	
3	0.64	-0.58	1.27	0.64	-0.84	-0.21	0.66	-0.17	2.01	-1.31	
平均值											$\bar{X}_{a=}$
极差											$\bar{R}_{a=}$
B 1	0.08	-0.47	1.19	0.01	-0.56	-0.2	0.47	-0.63	1.80	-1.68	
2	0.25	-1.22	0.94	1.03	-1.20	0.22	0.55	0.08	2.12	-1.62	
3	0.07	-0.68	1.34	0.2	-1.28	0.06	0.83	-0.34	2.19	-1.50	
平均值											$\bar{X}_{b=}$
极差											$\bar{R}_{b=}$
C 1	0.04	-1.38	0.88	0.14	-1.46	-0.29	0.02	-0.46	1.77	-1.49	
2	-0.11	-1.13	1.09	0.20	-1.07	-0.67	0.01	-0.56	1.45	-1.77	
3	-0.15	-0.96	0.67	0.11	-1.45	-1.49	0.21	-0.49	1.87	-2.16	
平均值											$\bar{X}_{c=}$
极差											$\bar{R}_{c=}$
零件 平均值											$\bar{\bar{X}} =$ $R_p =$
$\bar{\bar{R}} = (\bar{R}_{a=} + \bar{R}_{b=} + \bar{R}_{c=}) / [\text{评价人} =] =$											$\bar{\bar{R}} =$
$\bar{X}_{DIFE} = [\text{Max}\bar{X} =] - [\text{Min}\bar{X} =] =$											
$*UCL_R = [\bar{R} =] \times [D_4 =] =$											
<p>*2 次测量时 <math>D_4=3.27</math>, 3 次测量时 <math>D_4=2.58</math>。UCL<sub>R</sub> 代表了个别 R 值的限值。将那些超出限值的点圈出，查明原因并采取纠正措施。让相同的评价人使用相同的量具用原来的方法重新读值，或剔除这些数值并由其余的数值重新平均和计算 <math>\bar{\bar{R}}</math>，以及控制限值。</p>											

图 12: 量具重复性和再现性数据收集表

结果分析- 图示法<sup>48</sup>

使用图表工具是很重要的。使用哪种特定的图示取决于用于收集数据的实验设计。在进行其它的统计分析之前，应该使用图表工具对数据进行系统地筛选，从而找出变差的明显的特殊原因。

下面是一些经证明很有用的分析技术（参见方差分析法）

从测量系统分析中得到的数据可通过控制图画显示出来。通过使用控制图来回答与测量系统有关的问题，这一见解已被 Western Electric 所采用（见参考清单中 AT&T 统计质量控制手册）。

## 平均值图

以零件编号顺序画出由每个评价人对每个零件多次读值的平均值。该图可以用来确认评价人之间的一致性。

如果以极差的平均值计算所确定的总平均值和控制限制也画出，则这产生的平均值图可用来显示测量系统的“实用性”。

控制限以内区域表示测量的敏感性（干扰）。由于研究中使用的零件组代表过程变差，大约一半或一半以上的平均值应该落在控制限以外。如果数据显示出这种图形，则测量系统应该是适合进行检验出零件之间的变差，以及能为过程的控制提供有用的信息；如果少于一半的数据点落在并且测量系统能够提供过程分析和过程控制有用的信息。如果少于一半的均值落在控制限外边，则测量系统缺乏足够的分辨率或样本不能代表期望的过程变差。

图 13: 均值图-“层叠的”<sup>49</sup>

对图进行评价可知：测量系统有足够的解析度来测量样本零件所代表的过程变差。没有发现明显的评价人与评价人之间的差别。

图 14: 平均值图“非重迭画出”

---

<sup>48</sup> 这些分析的详细描述超出了本书的范围。如欲获得更多的信息，参见参考书并寻求由能力的统计资协助。

<sup>48</sup> 在 ANOVA 方法中，这也被称为评价人与零件相互作用图。

## 极差图

极差图被用来确定过程是否受控。原因是无论测量系统误差有多大，控制限将包含该误差。这就是为什么需要在进行相应的测量系统研究之前，需要识别并消除特殊原因变差的原因。

将由每个评价人对每个零件多次测量读值的极差，画在一个包括了极差平均值和控制限的标准极差图上。从被画在图上数据的分析，可以得到一些有用的解释。如果所有的极差均受控，则说明所有评价人都进行了相同的工作。

如果某个评价人是在控制限之外，则说明他使用的方法与其它人不一致。

如果所有评价人均有一些超出控制范围的点，则说明该测量系统对评价人的技巧较敏感，需要进行改进以获得有效的数据。

图表应该不是显示数据对于评价人或零件关系的图形

极差不是对数据的排序。不能像一般控制图趋势分析来使用，即使画出来的数据点是用线条连接的。

稳定性以一个点或多个超过控制限的点来确定，评价人之间或零件之间。稳定性分析要考虑到实用性和统计的显著性含义。

极差图可以帮助确定

- 与重复性有关的统计控制
- 测量过程中评价人之间对每个零件的一致性。

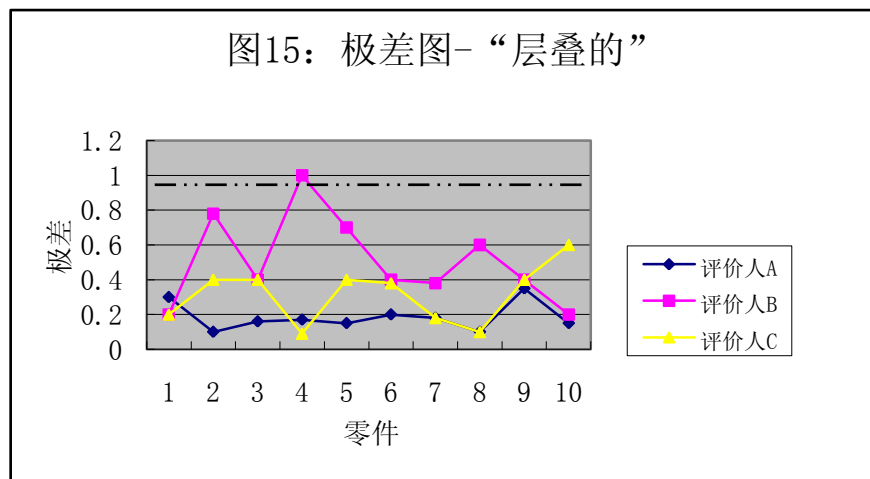


图 15：极差图-“重迭画出”

### 图 16: 极差图-“不层叠”

评审以上图表显示评价人的变差之间存在差异。

#### 链图

画出所有评价人对零件的所有读值以获得对下列的理解:

- 个别零件在变差一致性上的影响
- 奇异读值的呈现（即不正常的读值）

### 图 17: 零件的链图

对上图进行分析可知: 没有奇异数据或不一致的零件。

#### 散点图

将个别的读值依评价人所测量的零件绘制图表, 以获得下列的理解:

- 评价人之间的一致性
- 显示可能的分离
- 零件-评价人之间的相互作用

对图 18 进行分析可知, 没有指出任何明显的分离, 但指示出评价人 C 的读值可能比其他人小。

**图 18: 散点图**

### 震荡图

在震荡图中，依评价人所测量的零件画出读值的最高值、最低值以及平均值（见图 19），通过这图可理解：

- 评价人之间的一致性
- 分离的呈现
- 零件-评价人之间的相互作用

**图 19: 震荡图**

对图 19 进行分析可知，没有呈现任何明显的分离，但指出评价人 B 可能有较大的变差。

### 误差图

测量系统分析的数据可以通过“误差图”（见图 20）来被分析，即画出个别读值相对于可接受的参考值的误差。每个零件的个别读值偏差或误差可用以下公式计算：

误差=观察值-参考值

或

误差=观察值-零件的测量平均值

选用哪个公式取决于是否能够得到被测零件的参考值。



评价人 ■ A ■ B ■ C

图 20: 误差图

对上图进行评审可知:

- 评价人 A 具有整体性的正向偏倚
- 评价人 B 的变差最大, 但无明显的偏倚
- 评价人 C 有整体的负向偏倚。

### 正常化直方图

直方图 (图 21) 呈示参加这研究的评价人误差的频率分布图示法。这图也可显示所有评价人联合的频率分布。

如果可获得参考值, 则:

误差=观测值-参考值

否则: 正常化的数值=观察值-零件平均值

直方图能提供一快速目视观察这误差是如何分布的。甚至在对数据分析之前, 便能识别出评价人所进行的测量是否存在偏倚或不一致的问题。

对直方图进行分析 (图 21) 可以加强误差图分析的结论。它们同样指出只有评价人 B 呈现对称图形, 评价人 A 和 C 均由于偏倚而带来一系统上变差的来源。

**图 21： 正常化直方图<sup>50</sup>**

### 平均值 X—Y 图

由各评价人对每个零件多次读值的平均值，其与参考值或零件总平均值作为相对指数的图示法（见图 22），这图可以帮助确定：

- 线性（如果使用了参考值）
- 评价人之间线性的一致性

**图 22： 平均值 X-Y 图**

---

<sup>50</sup> 注意：每个直方图的“0.0”数值是与其它的直方图对齐。

## X-Y 比较图

由每个评价人对每个零件的多次读数的平均值,其与另一评价人的结果成为相对指数的图示法。本图是把一评价人获得的数据与其它的评价人数据进行比较(见图 23),如果评价人之间完全一致,则画出的点将形成一条与起源成 45 度角的直线。

图 23; X-Y 比较图

---

## 数值计算

量具重复性和再现性的计算如图 24 和 25 所示。图 24 是数据收集表格,记录了所有的研究结果。图 25 是报告表格,记录了所有识别信息和依据指定公式进行的所有计算。

本手册的表格范例部分提供了可复制的空白表格。在收集数据之后所进行的计算程序如下:

(以下参考图表 24)

1) 第 1, 2, 3 行中最大的值减去它们中的最小值,把结果记入第 5 行。在第 5, 6, 7 行和 11, 12, 13 行重复以上步骤,并将结果分别记录在第 10 行和第 15 行。

2) 填入第 5, 10, 15 行的数据是极差,所以它们都是正值。

3) 将第 5 行的数据相加,然后除以零件抽样数量即得到了第一个评价人试验的极差均值  $R_a$ ,对第 10 行和第 15 行进行同样的计算得到  $R_b$  和  $R_c$ 。

4) 将第 5, 10, 15 ( $R_a$ ,  $R_b$ ,  $R_c$ ) 行的平均值转到到第 17 行。用它们的和除以评价人数,结果记入 R 栏(所有极差的平均值)。

5) 将 R (平均值)输入到 19 和 20 行,乘以  $D_4^{51}$  得到 控制限上限和下限。注意如果进行 2 次测量,则  $D_4$  为 3.27。将该单独极差的上控制限 ( $UCL_R$ ) 记到第 19 行。

注意:对于试验少于 7 次的情况下,极差的下控制限 ( $LCL_R$ ) 为 0。

---

<sup>51</sup> 见统计过程控制参考手册, 1995, 或其它统计来源的因子表。

6) 对于极差结果大于  $UCL_R$  计算结果的任何读值, 让原来的评价人对原来的零件进行重新测量, 或者剔除那些值, 基于新的样本容量重新计算  $R$  和  $UCL_R$  值。纠正造成失控的特殊原因。如果用先前讨论过的控制图作图或分析数据, 这种情况已经被修正了, 在这里就不会出现。

7) 加总第 1, 2, 3, 6, 7, 8, 11, 12 和 13 行的读值, 然后将每行的总和除以抽样的数量, 并将结果填入该行最右边标有“平均值”的栏位中。

8) 将 1, 2, 3 的平均值相加, 用该总和除以测验次数, 将所有结果输入第 4 行  $X_a$  格中。第 6, 7, 8 和 11, 12, 13 行重复以上步骤, 将结果输入第 9, 14 行的相应  $X_b$ ,  $X_c$  格中。

9) 将第 4, 9, 14 行平均值 ( $X_a$ 、 $X_b$ 、 $X_c$ ) 中的最大值和最小值, 并将它们填入第 18 行的适当位置并求出它们的差; 将这差值填入第 18 行标有  $X_{DIFF}$  的空白处。

10) 计算每次测量每个零件的读值之和, 然后除以总测量次数, (试验次数乘以评价人人数)。将结果填在第 16 行为每个平均值提供的空白处。

11) 用零件平均值的最大值减去零件的平均值中的最小值, 并将结果填入到第 16 行中标示  $R_p$  格中。 $R_p$  表示零件不均值的极差。(以下步骤参考图 25)

12) 将计算的结果值  $\bar{R}$ ,  $X_{DIFF}$ ,  $R_p$  填入报告表格所提供的空白处。

13) 进行报告表格左侧标题为“测量单元分析”栏之下的计算。

14) 进行报告表格右侧标题为“总变差%”栏之下的计算。

15) 检查计算结果以确认没有错误。

评价人/试验#	零件										平均值
	1	2	3	4	5	6	7	8	9	10	
A 1	0.29	-0.56	1.34	0.47	-0.8	0.02	0.59	-0.31	2.26	-1.36	0.1
2	0.41	-0.68	1.17	0.5	-0.92	-0.11	0.75	-0.20	1.99	-1.25	0.1
3	0.64	-0.58	1.27	0.64	-0.84	-0.21	0.66	-0.17	2.01	-1.31	0.2
平均值	0.447	-0.607	1.260	0.537	-0.853	-0.100	0.667	-0.227	2.087	-1.307	$\bar{X}_a = 0.19$
极差	0.35	0.12	0.17	0.17	0.12	0.23	0.16	0.14	0.27	0.11	$\bar{R}_a = 0.1$
B 1	0.08	-0.47	1.19	0.01	-0.56	-0.2	0.47	-0.63	1.80	-1.68	0.0
2	0.25	-1.22	0.94	1.03	-1.20	0.22	0.55	0.08	2.12	-1.62	0.1
3	0.07	-0.68	1.34	0.2	-1.28	0.06	0.83	-0.34	2.19	-1.50	0.0
平均值	0.133	-0.790	1.157	0.413	-1.013	0.027	0.617	-0.297	2.037	-1.600	$\bar{X}_b = 0.06$
极差	0.18	0.75	0.40	1.02	0.72	0.42	0.36	0.71	0.39	0.18	$\bar{R}_b = 0.5$
C 1	0.04	-1.38	0.88	0.14	-1.46	-0.29	0.02	-0.46	1.77	-1.49	-0.2
2	-0.11	-1.13	1.09	0.20	-1.07	-0.67	0.01	-0.56	1.45	-1.77	-0.2
3	-0.15	-0.96	0.67	0.11	-1.45	-1.49	0.21	-0.49	1.87	-2.16	-0.2
平均值	0.073	-1.157	0.880	0.150	-1.327	-0.483	0.080	-0.503	1.697	-1.807	$\bar{X}_c = -0.25$
极差	0.19	0.42	0.42	0.09	0.39	0.38	0.20	0.10	0.42	0.67	$\bar{R}_c = 0.3$
零件平均值	0.169	-0.851	1.099	0.367	-1.064	-0.186	0.454	-0.342	1.940	-1.571	$\bar{X} =$ $R_p =$ 0.00 3.5
$([\bar{R}_a = 0.184] + [\bar{R}_b = 0.513] + [\bar{R}_c = 0.328]) / [\#评价人个数 = 3] =$											$\bar{R} = 0.34$
$[\text{Max}\bar{X} = 0.1903] - [\text{Min}\bar{X} = -0.2543] = \bar{X}_{DIFF} = 0.446$											
<ul style="list-style-type: none"> <li><math>[R = 0.3417] * [D_4 = 2.58] = UCL_R = 0.8816</math></li> </ul>											
<ul style="list-style-type: none"> <li>当测量次数为 2 次时 <math>D_4 = 3.27</math>, 3 次时 <math>D_4 = 2.58</math>。UCL<sub>R</sub> 表示 R 的界限。圈出那些超出界限值的点, 查明原因并采取纠正措施。让相同的评价人使用相同的量具用原来的方法重新读值, 或剔除这些数值并由其余数值再次平均和计算 R 以及控制限值。</li> </ul>											
注:											

图 24: 完成的 GR&R 数据收集表

量具重复性和再现性报告													
零件编号和名称:	量具名称:	日期:											
特性:	量具号:	操作人:											
规范:	量具类型:												
$\bar{R} = 0.3417$	$X_{DIFF} = 0.4446$	$R_p = 3.511$											
测量单元分析		% 总变差 (TV)											
重复性—设备变差 (EV) $EV = \bar{R} \times K_1$ $= 0.3417 \times 0.5908$ $= 0.20188$		<table border="1" style="margin-left: auto; margin-right: auto;"> <tr><th>试验</th><th><math>K_1</math></th></tr> <tr><td>2</td><td>0.8862</td></tr> <tr><td>3</td><td>0.5908</td></tr> </table>	试验	$K_1$	2	0.8862	3	0.5908	$\%EV = 100[EV/TV]$ $= 100[\frac{0.20188}{1.14610}]$ $= 17.62\%$				
试验	$K_1$												
2	0.8862												
3	0.5908												
再现性—设备变差 (AV) $AV = \sqrt{(\bar{X}_{DIFF} \times K_2)^2 - (EV^2 / (nr))}$ $= \sqrt{(0.4446 \times 0.523)^2 - (0.20188^2 / (10 \times 3))}$ $= 0.22963$		<table border="1" style="margin-left: auto; margin-right: auto;"> <tr><th>评价人</th><th>2</th><th>3</th></tr> <tr><td><math>K_2</math></td><td>0.7071</td><td>0.5231</td></tr> </table>	评价人	2	3	$K_2$	0.7071	0.5231	$\%AV = 100[AV/TV]$ $= 100[\frac{0.22963}{1.14610}]$ $= 20.04\%$				
评价人	2	3											
$K_2$	0.7071	0.5231											
重复性和再现性(GRR) $GRR = \sqrt{EV^2 + AV^2}$ $= \sqrt{(0.20188^2 + 0.22963^2)}$ $= 0.30575$		<table border="1" style="margin-left: auto; margin-right: auto;"> <tr><th>零件</th><th><math>K_3</math></th></tr> <tr><td>2</td><td>0.7071</td></tr> <tr><td>3</td><td>0.5231</td></tr> <tr><td>4</td><td>0.4467</td></tr> <tr><td>5</td><td>0.4030</td></tr> </table>	零件	$K_3$	2	0.7071	3	0.5231	4	0.4467	5	0.4030	$\%GRR = 100[GRR/TV]$ $= 100[\frac{0.30575}{1.14610}]$ $= 26.68\%$
零件	$K_3$												
2	0.7071												
3	0.5231												
4	0.4467												
5	0.4030												
零件变差(PV) $PV = R_p \times K_3$ $= 1.10456 \times \frac{1.10456}{3.511}$ $= 1.10456$		<table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>3</td><td>0.5231</td></tr> <tr><td>4</td><td>0.4467</td></tr> <tr><td>5</td><td>0.4030</td></tr> </table>	3	0.5231	4	0.4467	5	0.4030	$\%PV = 100[PV/TV]$ $= 100[\frac{1.10456}{1.14610}]$ $= 96.38\%$				
3	0.5231												
4	0.4467												
5	0.4030												
总变差(TV) $TV = \sqrt{GRR^2 + PV^2}$ $= \sqrt{(0.30575^2 + 1.10456^2)}$ $= 1.14610$		<table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>6</td><td>0.3742</td></tr> <tr><td>7</td><td>0.3534</td></tr> <tr><td>8</td><td>0.3375</td></tr> <tr><td>9</td><td>0.3249</td></tr> <tr><td>10</td><td>0.3146</td></tr> </table>	6	0.3742	7	0.3534	8	0.3375	9	0.3249	10	0.3146	$ndc = 1.41(\frac{PV}{GRR})$ $= 1.41(\frac{1.10456}{0.30575})$ $= 5.094 \dots 5$
6	0.3742												
7	0.3534												
8	0.3375												
9	0.3249												
10	0.3146												
表中所有的有关理论和常数的资料参见 MSA 参考手册第三版													

图 25: 量具 GR&R 报告

## 结果分析---数值法

图 24 和 25 的量具重复性和再现性数据收集表和报告表,提供了研究数据的数值分析<sup>52</sup>。这种分析将可以估计变差、整个测量系统变差占过程变差的百分比,以及测量系统的重复性、再现性和零件的变差,这些信息需要与图示分析法所得的全部结果进行比较。

在报告表(图 25)左侧的测量单元分析的下面,是对每个变差组成部分的标准差计算。

重复性或设备变差(EV 或  $\sigma_E$ )由极差平均值 R 乘以一个常数( $K_1$ )来决定。 $K_1$ 取决于量具研究中的测量次数,其值为从附录 C 查到的  $d_2^*$ 的倒数,  $d_2^*$ 取决于测量次数(m)和零件的数量乘以评价人数(g)(为了计算  $K_1$ 值,这值假设 g 大于 15)。

再现性或评价人变差(AV 或  $\sigma_A$ )由评价人平均值的最大差值( $X_{DIFF}$ )乘以一个常数  $K_2$ 来决定。 $K_2$ 取决于量具研究中的评价人的人数,其值为从附录 C 中查到的  $d_2^*$ 的倒数,  $d_2^*$ 取决于评价人的人数(m)且  $g=1$ ,因为只有一个极差计算。由于评价人变差包含在设备变差中,因此必须通过减去设备变差的一个分数来对其进行调整。因此评价人变差(AV)可由下列计算得到:

$$AV = \sqrt{\frac{(X_{DIFF} \times K_2)^2 - (EV)^2}{nr}}$$

这里, n=零件数, r=测量次数

如果根号下的值为负,评价人的变差(AV)为 0。

测量系统变差的重复性和再现性(GRR 或  $\sigma_M$ )的计算为设备变差的平方加上评价人变差的平方,然后再开根号,如下式:

$$GR \& R = \sqrt{(EV)^2 + (AV)^2}$$

---

<sup>52</sup> 示例中的数据结果如是由人工计算,结果被圆整到小数点后一位。由计算机程序计算则中间值应保留到计算机/程序语言的最大精度。从一个有效的计算机程序获得的结果可以与示例中结果的第二个小数位或更后面的小数位有区别,但最终分析将保持一致。

零件的变差 (PV 或  $\sigma_p$ ) 是由零件平均值的极差 ( $R_p$ ) 乘以一个常数 ( $K_3$ ) 确定。  $K_3$  取决于量具研究中使用零件的数量, 其值为从附录 C 查到的  $d_2^*$  的倒数。  $d_2^*$  取决于零件的数量 ( $m$ ) 和 ( $g$ ), 在这情况下,  $g=1$ , 因为只有一个极差计算。 这研究的总变差 (TV 或  $\sigma_T$ ), 是加总了重复性和再现性变差的平方与零件的变差 (PV) 的平方, 再开根号计算而得到。

$$TV = \sqrt{(GRR)^2 + (PV)^2}$$

如果已知该过程变差, 并且其值以  $6\sigma$  为基础, 则可以用它来代替从量具研究数据中计算得到的总研究变差 (TV)。 也就是说可通过以下两个公式进行计算来完成:

$$1) \quad TV = \frac{\text{过程变差}}{6.00}$$

$$2) \quad PV = \sqrt{(TV)^2 - (GRR)^2}$$

以上两个值 (TV 和 PV) 可以用来替代前面计算的量。

一旦确定了量个研究中的各个因素的变差后, 可将它们与总变差 (TV) 进行比较。 也就是完成量具报告表 (图 25) 右侧的 “%总变差” 下方的计算。

设备变差 (%EV) 占总变差 (TV) 的百分比, 被计算为  $100[EV/TV]$ 。 其他因素占总变差的百分比计算方法相似, 如下:

$$\%AV = 100[AV/TV]$$

$$\%GRR = 100[GRR/TV]$$

$$\%PV = 100[PV/TV]$$

**每个因素所占的百分比的和将不等于 100%。**

需要对以上总变差的百分比结果进行评价, 以确定该测量系统对其预期的使用是否为可接受。

如果分析是以公差为基础来代替以过程变差为基础的话, 则可对量具重复性和再现性 GRR 报告表 (图 25) 进行修改, 使格式右边的总变差的百分比由公差的百分比来替代。 在这种情况下, %EV, %AV, %GRR 和 %PV 的计算公式中的分母是由公差除以 6 来代替总变差 (TV)。 无论使用哪一种或两种方法都用, 是取决于这测量系统的期望用途以及顾客的期望。

在这数据分析的最后一个步是确定区别分类数 (the number of distinct categories), 这能由该测量系统可靠地分辨, 这是可以覆盖预期的产品变差的非重迭 97% 置信区间<sup>53</sup>。

$$Ndc = 1.41(PV/GRR)$$

<sup>53</sup> ndc 与 “Wheeler” “的分级比率” 相同, 在中被定义, 第一版 wheeler 和 LydAy, 1984。 作为确定有效分辨率的一种可选择方法, 见 1989 年出版的本书第二版。



如果图示分析法没有发现特殊原因的变差，则可用第二章第 D 节的比例法则来确定量具和重复性和再现性（%GRR）。

另外，ndc 应该四舍五入取整数，且要能大于或等于 5。

### 方差分析法（ANOVA）

方差分析（ANOVA）是一种标准的统计技术，可用它来分析测量误差和一测量系统研究中的其它变差来源。在变差的分析中，变差可以分解为四类：零件、评价人、零件与评价人之间的交互作用，以及由于量具造成的重复误差。

与平均值和极差法相比，ANOVA 方法有以下优点：

- 有能力解决任何实验的作业准备
- 可以更精确地估计变差
- 可从实验数据中获得更多信息（如零件与评价人之间的相互作用影响）。

其缺点是数据计算更复杂，而且要求一定程度的统计学知识以解释它的结果。以下章节中 ANOVA 方法是一建议我方法，特别是可以使用电脑时。

### 随机化和统计独立性

数据收集的方法在 ANOVA 中很重要。如果数据不是以随机方式收集的，将会带来一偏倚值的误差来源。为了确保（n）个零件、（k）个评价人、（r）次测量的平衡设计，一个简单的方法就是通过随机化。一个随机化的通用方法是：在一张纸片上写 A1，代表第一个评价人对第一个零件的测量，写下 A（n）代表第一个评价人对第 n 个零件的测量；用同样的程序对下一个评价人进行编号，直到第 K 个评价人。类似的标记将被用在 B1，C1 表示第二和第三个人对第 1 个零件的测量。等所有的 nk 组合都被写好后，把这些纸条放入一个帽子或碗中；每次抽取一张纸条，这纸条的组合（A1，B2，…）就是这量具研究所进行的测量顺序。所有的 nk 组合的纸条均抽出来后，把他们再全都放回到帽子中，再重复以上程序；重复该程序 r 次以确定每次重复的实验顺序。

也可以使用其它方法来进行随机抽样。但应该注意实施随机、抽样、随意抽样与自由抽样<sup>54</sup>等的区别。

### 进行研究

可以与图 12 相似的表格，以随机地方式收集数据。在这个范例中，选取了 10 个零件和 3 个评价人，且这实验在实施了随机的顺序下对每个零件和每位评价人的组合测量 3 次。

---

<sup>54</sup> 见《评价测量过程》，wheeler 和 LydAy，第二版，1989，P27。

## 图示分析法

前面所讨论的任何图示法，可被使用在 ANOVA 研究中对收集数据的图形分析的部分。用这些方法可对数据进行确认并进一步的提供理解（如趋势、周期等）。

建议使用一种被称为*相互作用图* (*interaction plot*) 的图示法。这个图可以确认 F 试验对相互作用是否显著。在图 26 描绘的相互作用图，其横向轴是零件编号 (1, 2, ...等)，纵轴是每个评价人对每个零件的测量平均值。将每个评价人对每个零件测量平均值的点连起来可以得到 k 条线 (评价人数量)。该图作如下解释：如果 k 条线是平行的，则没有交互关系。若线不平行，则交互作用就是显著的。相交的角度越大，交互作用越大。应该采取适当的措施来消除形成交互作用的原因。在图 26 所示的范例中，线几乎平行，表明没有显著的相互作用。

**图 26：相互作用图**

另外一种图有时是有意思的—余数图。该图能更好地检查假设的有效性。这种假设是量具(误差)是从一个正态分布的一项随机变数。画在图上的余数是观测值与预测值之间有差值，预测值是每个评价人对每个重复读值的平均值。如果这些余数不是随机地散布在零(水平参考线)的上侧和下侧，则这可能是由于假设是错误的，建议对数据进行进一步的调查。

**图 27：余数图**

## 数值计算

虽然这些数值可以用人工计算,但许多人利用电脑程式来产生所谓的方差分析法(ANOVA)表格(见附录A)。

这里的 ANOVA 表格由 5 列所构成(参见表 8)

- 来源 列是造成变差的原因
- *DF* 列是关于这来源的自由度
- *SS* 或平方的总和 列是围绕来源平均值附近的偏差
- *MS* 或平均平方列为平方总和除以自由度
- *F-比率*列是一个计算值,用来确定来源数值的统计显著性

ANOVA 表用于将总变差分解成 4 个部分:零件、评价人、零件与评价人之间的相互作用,以及由于仪器造成的重复性。

为了便于分析,将负的变差要素设定为零。

如在平均值和极差分析法一样,这些讯息是用来确定测量系统的特性。

表 8 为从图 24 中的数据范例,用 ANOVA 计算的结果,假设是一个固定的影响模式。表 10 为 ANOVA 分析法与平均值和极差法的比较。表 11 是对 ANOVA 分析法中的 GRR 报告。

变差来源	DF	SS	MS	F
评价人	2	3.1673	0.58363	34.44
零件	9	88.3619	9.81799	213.52
评价人和零件	18	0.3590	0.01994	0.434
设备	60	2.7589	0.04598	
总计	89	94.6471		
*显著水平 $\alpha=0.05$				

表 8: ANOVA 表格

变数的估计值	标准差 ( $\sigma$ )	%总变差	贡献率%
$\tau^2 = 0.039973$ (重复性)	$EV=0.199933$	18.4	3.4
$\omega^2 = 0.051455$ (评价人)	$AV=0.226838$	20.9	4.4
$\gamma^2 = 0$ (相互作用)	$INT=0$	0	0
系统= $0.09143(\tau^2 + \omega^2 + \gamma^2)$	$GRR=0.302373$	27.9	7.8
$\sigma^2 = 1.086446$ (零件)	$PV=1.042327$	96.0	92.2
总变差	$TV=1.085$	100.0	

表 9: ANOVA 分析法,变差%和贡献率  
(变数的估计值是以没有相互作用的模型为基础)

$$ndc = 1.41 \left( \frac{1.04233}{.30237} \right) = 4.861 \cong 4$$

$$\text{总变差(TV)} = 100 \sqrt{(GRR)^2 + PV^2}$$

$$\% \text{总变差} = 100 \left( \frac{\sigma_{\text{要素}}}{\sigma_{\text{全部}}} \right)$$

$$\% \text{总变差} = 100 \left( \frac{\sigma_{\text{要素}}^2}{\sigma_{\text{全部}}^2} \right)$$

方法	下 90%置信限 <sup>55</sup>	标准偏差	上 90%置信限 <sup>55</sup>	%总变差
GRR*	0.175	0.202	0.240	17.6
EV	0.133	0.230	1.016	21.1
AV	----	na	----	aa
交互作用	0.266	0.306	0.363	26.7
GRR		1.104		96.4
PV	0.177	0.200	0.231	18.4
ANOVA	0.129	0.227	1.001	20.9
EV	----	0	----	0
AV	0.237	0.302	1.033	27.9
交互作用		1.042		96.0
GRR				
PV				

\*在均值极差法中,不能估计交互作用

表 10: ANOVA 分析法与平均值和极差法的比较

零件编号和名称:	量具名称:	日期:	
特性:	量具编号:	完成人:	
规范:	量具类型:		
	标准差	%总变差	贡献率
重复性 (EV)	0.200	18.4	3.4
再现性 (AV)	0.227	20.9	4.4
评价人与零件(INT)	0	0	0
GRR	0.302	27.9	7.9
零件(PV)	1.042	96.0	92.2
测量系统对于过程控制和分析是可接受的			
说明:	总变差 (TV) = 1.085		
公差 = 不适用			
分级数 (ndc) = 4			

表 11: ANOVA 方法的 GRR 报告

<sup>55</sup>CL=置信界限。

## GRR 研究的分析

平均值和极差法和 ANOVA 分析法两者均能提供关于测量系统或量具的变差信息。

例如，重复性大于再现性，原因可能是：

- 仪器需要维护
- 可能对量具进行重新设计，以获得更好的严格度
- 需要对量具的夹紧或固定装置进行改进
- 零件内变差的太大

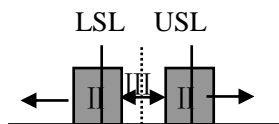
如果再现性比重复性大，则可能的原因是：

- 需要更好的对评价人进行如何使用和判读该量具仪器的培训
- 量具校准，刻度的不清晰

也许需要用某种形式的夹具帮助评价人更一致地使用量具。

### 第三章 — 第 C 节 计数型测量系统研究

计数型测量系统是一种测量数值为一有限的分类数量的测量系统。它与能获得一连串数值结果的计量型测量系统截然不同。通/止规是最常用的量具，它只有两种可能的结果；其它的计数型测量系统，例如目视标准，可能产生五到七个分类，如非常好、好、一般、差、非常差。在前面章节中所介绍的分析方法不能被用来评价这样的系统。



如第一章第 G 节所讨论，任何使用测量系统来作结论都会存在一定的风险；由于最大的风险位于分类的边界处，因此，最合适的分析方法是量化测量系统的变差成为量具性能曲线。

## 风险分析法

在一些计数情况下，不可能充分获得附有计量型参考数值的零件。在这种情况下，可使用下述方法来评价作出错误或不一致<sup>56</sup>决定的风险<sup>57</sup>：

- 假设性试验分析
- 信号探测理论

由于这些方法没有量化测量系统变差，应该只有在顾客同意的情况下才能使用。选择和使用于这样的技术应该基于一个良好的统计上的实践、以及了解一个错误决定对后续过程及最终顾客的影响。

计数型测量系统变差来源应该通过人的为因素和人机工程学研究的结果最小化。

## 可能的方法

### 情景

某生产过程处于统计受控状态，其性能指数为  $Pp=PpK=0.5$  是不可接受的。由于过程在生产不合格产品，于是被要求采取遏制措施以便从生产过程中挑出不可接受的产品。

<sup>56</sup> 这包括多位评价者的比较。

<sup>57</sup> 参见参考书目表。

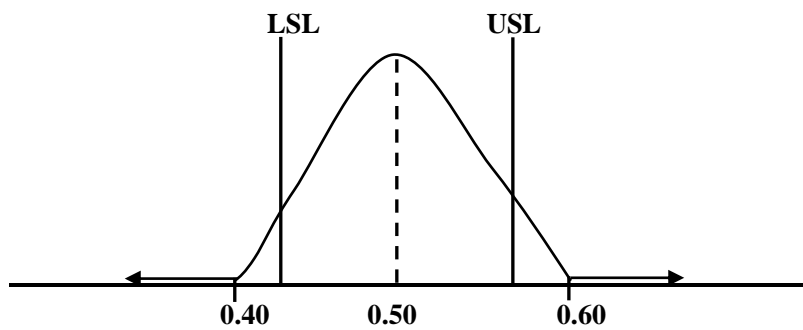


图 28：过程举例

具体的遏制措施是：过程小组采用一个计数型量具来对每个零件与一指定的限值进行比较，如果满足限值则可接受该零件，不满足的零件则拒收（如通/止量具）。许多这样的计数型量具基于一套基准零件来设定接受还是拒收；不像计量型量具，计数型量具不能显示一个零件多么好或多么坏，它只能指示该零件是被接受还是拒收。（即 2 个分类）。

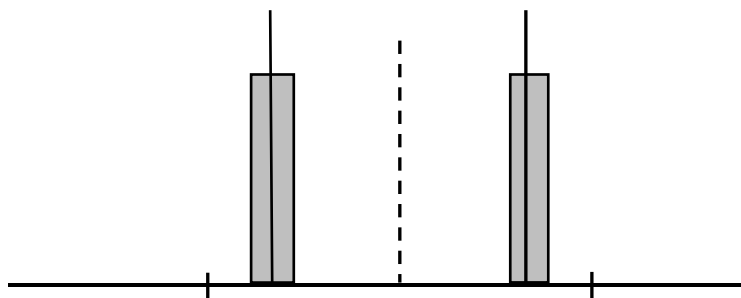


图 29：与测量系统有关的“灰色”区域

小组使用了一个%GRR 为公差<sup>58</sup>的 25%的特定量具。由于还没有被小组文件化，于是需要对这测量系统进行研究。小组已决定从过程中随机选取 50 个零件，以获得涵盖了整个过程范围的零件。使用三名评价人，每位评价人对每个零件评价三次。

<sup>58</sup> 过程由于变差比公差大，测量系统与公差进行比较而不是同过程变差进行比较是适当的。

零件	A-1	A-2	A-3	B-1	B-2	B-3	C-1	C-2	C-3	参考	参考值	代码
1	1	1	1	1	1	1	1	1	1	1	0.476901	+
2	1	1	1	1	1	1	1	1	1	1	0.509015	+
3	0	0	0	0	0	0	0	0	0	0	0.576459	-
4	0	0	0	0	0	0	0	0	0	0	0.566152	-
5	0	0	0	0	0	0	0	0	0	0	0.57036	-
6	1	1	0	1	1	0	1	0	0	1	0.544951	x
7	1	1	1	1	1	1	1	1	1	1	0.465454	x
8	1	1	1	1	1	1	1	1	1	1	0.502295	+
9	0	0	0	0	0	0	0	0	0	0	0.437817	-
10	1	1	1	1	1	1	1	1	1	1	0.515573	+
11	1	1	1	1	1	1	1	1	1	1	0.488905	+
12	0	0	0	0	0	0	0	0	0	0	0.559918	x
13	1	1	1	1	1	1	1	1	1	1	0.542704	+
14	1	1	0	1	1	1	1	0	0	1	0.454518	x
15	1	1	1	1	1	1	1	1	1	1	0.517377	+
16	1	1	1	1	1	1	1	1	1	1	0.531939	+
17	1	1	1	1	1	1	1	1	1	1	0.519694	+
18	1	1	1	1	1	1	1	1	1	1	0.484167	+
19	1	1	1	1	1	1	1	1	1	1	0.520496	+
20	1	1	1	1	1	1	1	1	1	1	0.477236	+
21	1	1	0	1	0	1	0	1	0	1	0.452310	x
22	0	0	1	0	1	0	1	0	1	0	0.545604	x
23	1	1	1	1	1	1	1	1	1	1	0.529065	+
24	1	1	1	1	1	1	1	1	1	1	0.514192	+
25	0	0	0	0	0	0	0	0	0	0	0.599581	-
26	0	1	0	0	0	0	0	0	1	0	0.547204	x
27	1	1	1	1	1	1	1	1	1	1	0.502436	+
28	1	1	1	1	1	1	1	1	1	1	0.521642	+
29	1	1	1	1	1	1	1	1	1	1	0.523754	+
30	0	0	0	0	0	1	0	0	0	0	0.561457	x
31	1	1	1	1	1	1	1	1	1	1	0.503091	+
32	1	1	1	1	1	1	1	1	1	1	0.505850	+
33	1	1	1	1	1	1	1	1	1	1	0.487613	+
34	0	0	1	0	0	1	0	1	1	0	0.449696	x
35	1	1	1	1	1	1	1	1	1	1	0.498698	+
36	1	1	0	1	1	1	1	0	1	1	0.543077	x
37	0	0	0	0	0	0	0	0	0	0	0.409238	-
38	1	1	1	1	1	1	1	1	1	1	0.488184	+
39	0	0	0	0	0	0	0	0	0	0	0.427687	-
40	1	1	1	1	1	1	1	1	1	1	0.501132	
41	1	1	1	1	1	1	1	1	1	1	0.513779	
42	0	0	0	0	0	0	0	0	0	0	0.566575	
43	1	0	1	1	1	1	1	1	0	1	0.462410	x
44	1	1	1	1	1	1	1	1	1	1	0.470832	+
45	0	0	0	0	0	0	0	0	0	0	0.412453	-
46	1	1	1	1	1	1	1	1	1	1	0.493441	+
47	1	1	1	1	1	1	1	1	1	1	0.486379	+
48	0	0	0	0	0	0	0	0	0	0	0.587893	-
49	1	1	1	1	1	1	1	1	1	1	0.483803	+
50	0	0	0	0	0	0	0	0	0	0	0.446697	-

表 12: 计数型研究数据



设定用 1 表示可接受的决定，0 为不可接受的决定。表 12 中所示的参考决定和计量参考值在一开始还没有被确定。表 12 还显示了“代码”列，分别用“-”“X”“+”代表零件是否位于 I 区，II 区及 III 区。

### 假设试验分析- 交叉表法

由于小组不知道零件的参考决定，他们展开了交叉表格来比较每个评价人与其它人的结果。

A 与 B 交叉表

		B		总计
		.00	1.00	
A	.00 计算	44	6	50
	期望的计算	15.7	34.3	50.0
	1.00 计算	3	97	100
	期望的计算	31.3	68.7	100.0
总计	计算	47	103	150
	期望的计算	47.0	103.0	150.0

B 与 C 交叉表

		C		总计
		.00	1.00	
B	.00 计算	42	5	47
	期望的计算	16.0	31.0	47.0
	1.00 计算	9	94	103
	期望的计算	35.0	68.0	103.0
总计	计算	51	99	150
	期望的计算	51.0	99.0	150.0

A 与 C 交叉表

		C		总计
		.00	1.00	
A	.00 计算	43	7	50
	期望的计算	17.0	33.0	50.0
	1.00 计算	8	92	100
	期望的计算	34.0	66.0	100.0
总计	计算	51	99	150
	期望的计算	51.0	99.0	150.0

这些表格的目的是确定评价人之间一致性的程度。为确定一致性的程度，小组使用了 kappa，这是用来衡量两个评价人对同一物体进行评价时，其评定结论的一致性。Kappa 为 1 时，表示有完全的一致性，为 0 时，表示一致性不比可能性来得好。Kappa 仅用于表格，表中两个变数有相同的分类值，且两个变数有相同的分类数量。

Kappa 是一种对评价人内部一致性的衡量，它测量在诊断区（获得相同评定的零件）中数量与那些基于可能性期望的数量是否有差别。

让  $P_o$  = 在对角栏框中，观测比例的总合

$P_e$  = 在角栏框中，期望比例的总合

则  $Kappa = (P_o - P_e) / (1 - P_e)$

**Kappa** 是一种程度而不是一种试验<sup>59</sup>。通过使用一种渐近的标准误差以形成一个  $t$  统计值来判断其大小。通用的比例法则是 **kappa** 值大于 0.75 则表示有很好的的一致性(**Kappa** 最大为 1); **kappa** 小于 0.4 则表示一致性差。

**Kappa** 不考虑评价人之间的不一致量有多大，只考虑他们之间是不是一致<sup>60</sup>。

通过评价人计算的 **Kappa** 程度，小组得到以下结论：

<b>Kappa</b>	<b>A</b>	<b>B</b>	<b>C</b>
<b>A</b>	—	<b>.86</b>	<b>.78</b>
<b>B</b>	<b>.86</b>	—	<b>.79</b>
<b>C</b>	<b>.78</b>	<b>.79</b>	—

这分析表明所有评价人与其它评价人之间有良好的—致性。

这种分析用来对确定评价人之间是否存在差异的需求，但不能告诉我们这测量系统从坏零件中挑选出好零件的能力。在本分析范例中，小组使用—计量型测量系统来评价零件，并应用其结果来确定其参考决定。

使用该新的信息建立了另一组交叉表，以便将每个评价人与参考决定进行比较。

A 与基准判断交叉表

		基准		总计
		.00	1.00	
A	.00 计算	45	5	50
	期望的计算	16.0	34.0	50.0
	1.00 计算	3	97	100
	期望的计算	32.0	68.0	100.0
总计	计算	48	102	150
	期望的计算	48.0	102.0	150.0

B 与基准判断交叉表

		基准		总计
		.00	1.00	
B	.00 计算	45	2	47
	期望的计算	15.0	32.0	47.0
	1.00 计算	3	100	103
	期望的计算	33.0	70.0	103.0
总计	计算	48	102	150
	期望的计算	48.0	102.0	150.0

C 与基准判断交叉表

		基准		总计
		.00	1.00	
A	.00 计算	42	9	51
	期望的计算	16.3	34.7	51.0
	1.00 计算	6	93	99
	期望的计算	31.7	67.3	99.0
总计	计算	48	102	150
	期望的计算	48.0	102.0	150.0

小组也计算了 Kappa 值以确定每个评价人与基准判断一致的程度。

	A	B	C
Kappa	.88	.92	.77

这些值可以被解释为每个评价人与基准有好的一致性。

然后，过程小组计算了测量系统的有效性。

有效性 = 作出正确决定的次数 / 总决定次数

<sup>59</sup> 如同所有这些分极评价，覆盖可能范围的大量的零件是必须的。

<sup>59</sup> 当观察是以一种顺序分级比例被测量时，为了更好地测量一致性，可以使用一种加权的 kappa。两个评价人之间的一致被作为 kappa 对待，不一致由分级的数量来测量，以此确定评价人的不一致。

变差来源	%评价人 <sup>1</sup>			得分与计数 <sup>2</sup>		
	评价人 A	评价人 B	评价人 C	评价人 A	评价人 B	评价人 C
总检查数	50	50	50	50	50	50
相配数	42	45	40	42	45	40
错误的拒收（由于评价人偏倚造成的拒收）				0	0	0
错误的接受（由于评价人偏倚造成的接受）				0	0	0
不相配				8	5	10
95%UCI	93%	97%	90%	93%	97%	90%
计算所得结果	84%	90%	80%	84%	90%	80%
95%LCI	71%	78%	66%	71%	78%	66%

	系统有效得分 <sup>3</sup>		系统有效得分与计数 <sup>4</sup>	
检查总数	50		50	
一致的数量	39		39	
95%UCI	6		64%	
计算所得结果	4%			
95%LCI	78%		78%	
	89%		89%	

注意：

- 1) 在所有测量中，评价人本身是一致的
- 2) 评价人对所有测量与已知的标准一致
- 3) 所有评价人本身与其它人之间是一致的
- 4) 所有评价人本身与其它人之间一致，并与参考值一致。
- 5) UCI 和 LCI 分别为置信区间边界的上限和下限

每对评价之间多次测量的假设可用零假设来表示：

$H_0$ ：两个评价人一致的有效性。

由于每个评价人结果的计算值均落在其它人的置信度区间内，小组决定不能拒绝零假设。这结论进一步证实了 KAPPA 测量得到的结论。

为进一步分析，一名组员列出一下列的表格，为每个评价人的结果提供指南：

决定 测量系统	有效性	错误率	错误警报率
评价人可接受的条件	≥90%	≤2%	≤5%
评价人可接受的边缘-可能 需改进	≥80%	≤5%	≤10%
评价人不可接受-需改进	<80%	>5%	>10%

对他们所已经得到的所有信息进行汇总，小组得到以下结论：

	有效性	错误率	错误警报率
A	84%	6.3%	4.9%
B	90%	6.3%	2.0%
C	80%	12.5%	8.8%

基于这信息，小组得到结论：当考量有了有效性、错误率与错误警报率，该测量系统就评价人 B 来说是有限度的接受，小组判断测量系统中评价人 B 可接受，评价人 A 处在边缘，评价人 C 不可接受。

尽管这些结论看起来是否定了他们在前面对于评价人之间没有统计上差别的结论，小组决定采用这个前面的结论，因为他们已经厌烦了所有这些分析，并且这些结论至少已证明了他们所提出的过程情况，对顾客来说也是可被接受的。

#### 关注点：

- 1) 关于可接受的风险，并没有以理论为基础的决策准则。以上指南是探索性的，并且是基于怎样才是“可接受”的个别“信念”下所发展的。最终的决策准则应该取决于对后续过程和最终顾客的影响（如风险）。这是一个客观事物的决定----而不是统计上的决定。
- 2) 上述分析是以数据为依据的。例如，如果过程能力指数为  $Pp=Ppk=1.00$ ，则所有的结论将可能正确，因为没有零件会落在测量系统的 II 区（灰色区）。

**图 30：具有  $Pp=Ppk=1.00$  的过程范例**

在这新的情况下，可以得出这样的结论：所有的评价人都是可接受的，因为将不会有决定误差。

3) 对于交叉表结果的意义通常有一个误解。例如, 评价人 B 的结果为例:

**B 与参考交叉表**

		基准		总计
		.00	1.00	
B	.00	数量 45	2	47
		在参考值内的% 93.8%	20%	31.3%
	1.00	数量 3	100	103
		在参考值内的% 6.3%	98%	68.7%
总计		数量 48	102	150
		在参考值内的% 100%	100.0%	100.0%

由于检验的目的是找出所有的不合格零件, 许多人视左上处为一个测量找到坏零件的有效性。这百分比表示将已经是坏的零件判定为坏零件的可能性:

$$\Pr(\text{称为坏零件} \mid \text{一个坏零件})$$

假设过程已经被改进  $P_p = P_{pk} = 1.00$ , 生产者关心的概率是:

$$\Pr(\text{零件是坏的} \mid \text{被称为坏的})$$

从上面的数据中确定以上结果, 必须应用贝叶斯理论。

$$\Pr(\text{不合格} \mid \text{判不合格}) = \frac{\Pr(\text{判不合格} \mid \text{判不合格}) * \Pr(\text{不合格})}{\Pr(\text{判不合格} \mid \text{判不合格}) + \Pr(\text{判不合格} \mid \text{判合格}) * \Pr(\text{不合格})}$$

$$\Pr(\text{不合格} \mid \text{判不合格}) = \frac{.938 * (.0027)}{.938 * (.0027) + .020 * (.9973)}$$

$$\Pr(\text{不合格} \mid \text{判不合格}) = .11$$

也就是说, 以上这些结果其零件被判定成是坏的, 实际上它只有十分之一的可能是坏的。

5) 这种分析不必使用计量型数据, 即使参考决定值已被确定且可获得时, 也不需要安排这相关的资料。

### 信号检查法

另一种替代的方法—信号检查理论来确定区域 II 的近似宽度, 从而确定测量系统的 GRR。

让  $d_i$  = 被所有评价人接受的最后一个零件与被所有评价人拒绝的第一个零件 (对于每个规范) 之间的距离。



## 分析方法<sup>62</sup>

就像对于任何测量系统一样,应该验证过程的稳定性,如有需要,还应进行监控。对于计数型测量系统,通常采用恒定样本的长期计数型控制图来验证其稳定性<sup>63</sup>。

对于计数型测量系统,量具性能曲线的概念(见第四章第 C 节)是被用来展开一个测量系统研究,以评估测量系统重复性和偏倚的数值,这分析可用于单边限值和双边限值的测量系统。对于双边限值测量系统,只有一个限值需要用来检查差异的线性和一致性的假设。为了方便起见,用下限进行讨论。

---

<sup>62</sup> 经允许引自“计数型量个系统分析”, J.McCaslin 和 G.GrusrRa,ASQC,1976。

<sup>63</sup> 注意:  $np > 4$ 。



通常,计数型测量系统研究由获得了多个被选定零件的参考值所构成。应该记录每个零件被评价的次数 (m),以及接受的总数 (a),从这些结果中,可以评估重复性和偏倚。

计数型研究的第一步是选取零件。研究中的每个零件的基准值已知是重要的。按近似等距离间隔选择 8 个零件。最大和最小值应该代表过程的范围。尽管这种选择不会影响结果的置信度,但是影响完成量个研究需要的零件总数量。这 8 个零件都要用量个测量, m=20 次,记录接受的数量 (a)。

为整体的研究,最小的零件必须 a=0;最大的零件 a=20;其它 6 个零件则在  $1 \leq a \leq 19$ 。如果这些准则不能被满足,需要选择更多的已知基准值 (X) 的零件,用量个测量,直到上述条件得到满足。如果最小的值  $a \neq 0$ ,则选取更小的零件并评价直到 a=0。如果最大的值  $a \neq 20$ ,则选取更大的零件直到 a=20。如果其他 6 个零件不满足  $1 \leq a \leq 19$ ,则在全范围内选取额外的零件,可以选择早已在研究中被测零件在中间的那些点。在 a=0 端,第一间隔起始于 a=0 的最大测量处; a=20 端,第一个间隔起始于 a=20 的最小测量处。为了获得更好的结果,抽样应该在 a=0 和 a=20 两端进行,然后朝零件范围的中间选取。必要时可以重复上述程序直到满足准则要求。

一旦满足了数据收集准则后,必须按以下公式计算各个零件的接收概率:

$$P_a = \begin{cases} \frac{A+0.5}{m} & \text{若 } \frac{a}{m} < 0.5, & a \neq 0 \\ \frac{a-0.5}{m} & \text{若 } \frac{a}{m} > 0.5, & a \neq 20 \\ 0.5 & \text{若 } \frac{a}{m} = 0.5 \end{cases}$$

以上调整覆盖了  $1 \leq a \leq 19$  的情况。对这情况,当 a=0 时,设  $P_a=0$ ,除去 a=0 的最大参考值,此时其  $P_a=0.025$ ;当 a=20 时  $P_a=1$ ,除去 a=20 的最小参考值,此时其  $P_a=0.975$ 。

一旦每个  $X_T$  的  $P_a$  计算出来,就可以建立量具性能曲线 (GPC)。尽管 GPC 可用图形来表示 (见图 32),用正态概率图 (见图 31) 可以产生更精确的重复性和偏倚估计值。

将计算得到的概率绘制在正态概率图上,画出经过这些点的一条最适合直线。偏倚等于下限值减去  $P_a=0.5$  对应的测量参考值

或

$$\text{偏倚} = \text{规范下限值} - X_T (P_a=0.5 \text{ 时})$$

发现  $P_a=0.995$  和  $P_a=0.005$  之间对应的测量参考值的差异再除以  $1.08^{64}$  的调整系数即为重复性。

$$\text{重复性} = \frac{X_T(P_a=0.995 \text{ 处}) - X_T(P_a=0.005 \text{ 处})}{1.08}$$

为了确定偏倚是否明显偏离零, 采用下述统计量:

$$t = \frac{31.3 \times |\text{偏倚}|}{\text{重复性}}$$

如果该计算值明显大于  $2.093(t_{0.25, 19})$ , 则这偏倚是显著的偏离零。

下面将用一个范例来阐述计数型研究的数据收集和重复性、偏倚的计算。  
显著性是在  $\alpha=0.05$  水准

#### 范例:

用一个计数型的量具测量公差为  $\pm 0.010$  的尺寸。这量具是一个线末端 100% 自动检查量具, 它已受重复性和偏倚的影响。为完成计数研究, 用该量具测量 8 个零件, 每个零件各测量 20 次, 这 8 个零件的参考值为从 -0.016 至 -0.002, 间隔为 0.002。各零件接受次数为:

$X_T$	a
-0.016	0
-0.014	3
-0.012	8
-0.010	20
-0.008	20
-0.006	20
-0.004	20
-0.002	20

由于只有 2 个参考值位于  $1 \leq a \leq 19$ , 必须至少应再找 4 个零件。因此, 需要测量那些参考值在现有间隔的中间处的零件, 这些参考值及接受次数为:

-0.015	1
-0.013	5
-0.011	16

现在我们有 5 个参考值位于  $1 \leq a \leq 19$ 。程序要求再找到一个零件可以位于  $1 \leq a \leq 19$ 。为此, 评价了以下零件:

-0.0105	18
---------	----

<sup>64</sup> 1.08 调整因子由 20 个样本容量规定, 通过这种途径模拟确定。

现在已经满足了数据收集准则，采用二项调整可以计算它们的接受概率，如下所示：

$X_T$	a	$P_a$
-0.016	0	0.025
-0.015	1	0.075
-0.014	3	0.175
-0.013	5	0.275
-0.012	8	0.425
-0.011	16	0.775
-0.0105	18	0.875
-0.010	20	0.975
-0.008	20	1.000

将这些概率画在正态概率图上,如图 31 所示。画出一条经过这些点的最佳适合直线,可确定偏倚和重复性。偏倚等于对应的下限值减去  $P_a=0.5$  对应的测量参考值。

由图 31, 可得到:

$$\text{偏倚} = -0.010 - (-0.0123) = 0.0023$$

重复性等于发现  $P_a=0.995$  和  $P_a=0.005$  对应的测量参考值之差值,再除以 1.08。从图 31 可以得到:

$$R = \frac{|-0.0163| - |-0.0084|}{1.08} = \frac{0.0079}{1.08} = 0.0073$$

为了确定偏倚是否明显偏离零,计算:

$$t = \frac{31.3 \times |\text{偏倚}|}{R} = \frac{31.3 \times |0.0023|}{0.0073} = 9.86$$

象第五章图 C 所示的计量型量具性能曲线一样,计数型 GPC 也可绘制在空白图纸上(见图 32)。这可用两种方法中任何一种方法来完成。一种方法将可以在另一个规范限值处进行抽样研究。本例中,计数型研究的大样法也必须在规范上限处进行研究,并画出相应的计算值。

但是,基于前述假设,将不需要再次进行研究,因为在上限处的曲线形状应该与在下限值的曲线成“对称影像”,唯一需要考虑的是对应于  $X_T$  值的曲线位置,该位置由偏倚来确定。

曲线正确位置应被定义在  $P_a=0.5$  的点，且  $X_T$  值等于规范限值减去偏倚。本例中，该点  $X_T=0.010-0.0023=0.0077$ 。按这种方法绘制的 GPC 如图 31 所示。

重复性 = **0.0079**  
无调整

偏倚 = **0.010 - (-0.0123) = 0.0023**

**图 31：绘制在正态概率纸上的计数型量具性能曲线**

被测零件的基准值  
偏倚=0.0023      GRR=0.0079

**图 32：计数型量具性能曲线**

## 第四章

### 其它测量概念和实践

## 第四章- 第 A 节 实施复杂的或不可重复的测量系统

### 引言

本参考手册的重点是对每个零件的读值能被重复的读数的测量系统，但并不是所有的测量系统都有这种特性，例如：

- 破坏性测量系统
- 零件随着使用/试验会有变更的系统；例如，发动机或变速箱动力计试验

下面是一些测量系统分析方法的例子，包括手册中前面未被讨论的那些系统。这里不打算全面列出覆盖各种型式的测量系统，而只是给出不同的方法的一些例子。如果本手册不适用于你所具有的测量系统，建议你求助于合适的统计资源。

不可重复的测量系统	
<b>情景——非破坏性测量系统</b>	<b>范例</b>
在测量过程中零件不会被改变；即测量系统是非破坏性的并可使用具有下列属性零件（样本）： <ul style="list-style-type: none"> <li>• 静态的属性，或</li> <li>• 已经稳定的动态（变化）属性。</li> </ul>	<ul style="list-style-type: none"> <li>● 用于不是新的车辆/动力系统进行测功试验</li> <li>● 有计量型数据的泄漏试验</li> </ul>
已知特性（属性）的定性期限，并能延续至超过期望的研究期间，即经过一段预期的使用后，被测特性不会改变。	从单一材料批次中抽取样本进行质量的光谱测量
<b>情景——破坏性测量系统</b>	<b>示例</b>
试验台	<ul style="list-style-type: none"> <li>● 生产线末端 √发动机试验台 √变速箱试验台 √整车动力试验机</li> <li>● 有定性数据的泄露试验</li> <li>● 盐雾试验/湿度试验</li> <li>● 重力计</li> </ul>
其它不可重复的测量系统	<ul style="list-style-type: none"> <li>● 采用不允许重复的自动线上测量系统</li> <li>● 破坏性焊接测试</li> <li>● 破坏性电镀测试</li> </ul>

表 13：测量系统示例



本章中所描述的研究方法及不同的情况如下表所示：

稳定性研究					
情景	S1	S2	S3	S4	S5
在测量过程中零件不会被改变，即测量系统是非破坏性的（可重复），并且使用具有如下特性的零件（样本）： • 静态的属性，或 • 已经稳定的动态（变化）属性	√	√			
已经特性（属性）的定性期限，并能延续至超过期望的研究期间---经过一段预期使用后，测量特性不改变。	√	√			
破坏性测量系统			√	√	
不可重复的测量系统			√	√	
试验台					√

变差研究									
情景	V1	V2	V3	V4	V5	V6	V7	V8	V9
在测量过程 的零件不会被改变；即测量系统是非破坏性的（可重复），并且使用具有下列特性的零件（样本）： • 静态的属性，或 • 已经稳定的动态（变化）属性。	√								
上述情况的 P≥2 仪器		√							
破坏性测量系统			√	√					
不可重复的测量系统			√	√					
具有动态特性的测量系统；例如，试验台			√	√	√	√	√	√	
上述情况的 P≥3 仪器									√

表 14：基于不同类型的测量系统使用的方法

## 第四章- 第 B 节 稳定性研究

---

### S1: 单个零件<sup>65</sup>, 每次实施单一测量 应用:

a) 零件不会由测量过程而被改变的测量系统, 即测量系统是非破坏性的, 并可使用具有下列属性的零件:

- √ 静态特性, 或
- √ 已经稳定的动态(变化)属性。

b) 该特性(属性)的定性期限, 并延续超过预期的研究期间; 就是说, 在预期的使用期间, 被测特性不发生改变。

#### 假设:

- 已知(已有文件说明)该测量系统在预期的特性(属性)范围内, 具有一线性的回应
- 零件(样本)涵盖该特性的过程变差的预期范围。

#### 采用 X&mR 图表分析:

- 确定测量系统的稳定性
  - √ 将描绘的点和控制限值进行比较;
  - √ 寻找趋势(仅 X 图)。
- 比较  $\sigma_e = R/d_2^*$  (总测量误差) 和由变差研究(见下一章)得到的重复性估计值  $\sigma_E$  来进行比较。

- 如果已知参考值, 则确定偏倚:

$$\text{偏倚} = \bar{X} - \text{参考值}$$

---

<sup>65</sup> 如果对于过程是适用的可以使用一参考标准。

**S2:  $n \geq 3$  个零件<sup>66</sup>, 每次每个零件实施单一测量应用:**

a) 零件不会由测量过程而被改变的测量系统, 即测量系统是非破坏性的, 并可使用具有下列属性的零件:

- √ 静态特性, 或
- √ 已经稳定的动态(变化)属性。

b) 已知该特性(属性)的定性期限, 并延续超过预期的研究期间, 即在预期的使用期间中, 被测特性不发生改变。

**假设:**

- 已知(已有文件说明)这测量系统在预期的特性(属性)范围内, 具有一线性回应。
- 零件(样本)覆盖该特性过程变差的预期范围。

采用[Z, R]图分析, 此处  $z_i = x_i - \mu_i$

且  $\mu_i$  是参考标准值, 或者由对零件(样本)大量连续的读值平均值所确定的值。

- 确定测量系统的稳定性

- √ 将描绘的点和控制限值进行比较;
- √ 寻找趋势(仅 Z 图)。

- 将  $\sigma_e = R/d_2^*$  和由变差研究所得出的重复性估计  $\sigma_E$  进行比较<sup>67</sup>。

- 若已知参考值则确定偏倚:

$$\text{偏倚} = \bar{X} - \text{参考值}$$

- 若采用了  $n \geq 3$  个零件, 则确定线性。

- √ 零件(样本)必须涵盖特性的预期范围。
- √ 各零件(样本)应该被单独分析其偏倚和重复性。
- √ 使用第三章, 第 B 节所讨论的线性分析法来量化线性。

如果在本研究中, 使用了一个以上的仪器, 要确定这些装置的一致性(变差的同质性), 例如采用 F 试验, Bartlett 试验, Levene 试验<sup>68</sup>等。

<sup>66</sup> 如果对过程适用, 可以使用一参考标准。

<sup>68</sup> dataplot, national institute of standards and technology, statistical engineering division ([www.itl.nist.gov](http://www.itl.nist.gov))

### S3:从稳定过程中的大量样本

#### 应用:

测量系统必须同同质的独立可识别分布 (independent identically distributed ,iid) 样本(收集并隔离)来评价。各个零件 (样本) 的测量不被重复进行, 因此可以用破坏性的和不可重复的测量系统进行研究。

#### 假设:

- 已知特性 (属性) 的定性期限并延续超过预期的研究期间, 即在预期的使用和/或储存期间被测特性不改变;
- 这些零件 (样本) 覆盖该特性 (属性) 的过程变差的预期范围;
- 已知 (已由文件说明) 测量系统的线性涵盖了特性 (属性) 的预期范围。(如果这回应是非线性的, 则必须相应地校正读值)。

#### 进行分析:

- 通过对  $n \geq 30$  个零件的能力研究确定, 以确定总变差 (这种初步研究也应该被用来验证样本的一致性, 即所有零件 (样本) 来自单边形式的分布)。
- $\sigma^2_{\text{总}} = \sigma^2_{\text{过程}} + \sigma^2_{\text{测量系统}}$
- 测量从每个时间周期的隔离样本中的一个或多个零件, 使用带有控制限的 X&R 或 X&mR 图, 这控制限值由能力研究所确定。
- 将描绘点和控制限进行比较;
- 寻找趋势。
- 因为这些零件 (样本) 不会变化 (一个隔离的样本), 任何不稳定性迹象将归因于该测量系统的变化。

### S4: 分割样本 (通用), 每循环单一样本

#### 应用:

各个零件(样本)的测量部分不被重复, 故可以采用破坏性或不可重复的测量系统进行研究。

#### 假设:

- 特性 (性质) 的保存寿命已知并延续超过预期的研究时期, 即在整个预期的使用和/或储存期间被测特性不改变;
- 零件 (样本) 覆盖该特性 (属性) 的过程变差的预期范围;

- 这些样本被分成  $m$  部分。 $m=2$  部分时常称为试验一再试验研究。

用以下方法分析：

- 极差图，以便追踪测量的一致性（推翻“批量内部”一致性）。
- 将  $\sigma_e = R/d_2^*$  和由变差分析得到的重复性估计  $\sigma_E$  进行比较。
- 这是上边界的研究： $\sigma_e^2 = \sigma^2_{\text{重复性估计}} + \sigma^2_{\text{样本之间}}$
- 用图表以便追踪生产过程的一致性

S4 用于从不同批次中采用连续的（同质的）配对零件—S4a

该研究同 S4 一样，但是是从不同批次所抽取的同质性的零件。

它是一种上边界研究： $\sigma_e^2 = \sigma^2_{\text{重复性估计}} + \sigma^2_{\text{样本之间}} + \sigma^2_{\text{批次}}$

## S5: 试验台

在这种情况下，由多台测量仪器（试验台）评价连续生产产品的同一特性。这些产品是随机的分配到各试验台。

### S5a: 计数型数据回应

用 P 图进行分析：

- 确定试验台之间（判断的）一致性：使用包括各试验台结果的一张图表；
- 确定各试验台内的稳定性：对每个试验台使用单独的图表。

采用 p&mR 图分析整个系统的稳定性,这里 p 是在某个设定工作日内所有试验台的平均值。

### S5b: 计量型数据回应

采用方差分析（ANOVA）和图示技术分析<sup>69</sup>

- 以时间周期计算每个试验台（依特性）的 X&s。
- 确定试验台之间的一致性：一张包括所有试验台结果的 X&s 图表。
- 确定各个试验台内的稳定性：每个试验台单独的 X&s 图。
- 对试验台之间的一致性（变差的同质性）：例如，采用 F 试验、Bartlett 试验、Levene 试验；

<sup>69</sup> 参见 James 博士的“量个 R&R 数据图形表示”ASQC 学报,ASQC,1991。

- 通过与比较试验台平均值的比较，确定所有试验台是否在同一目标上；例如采用一单向 ANOVA 分析。如果存在任何差别，通过采用例如 Tukey 的 T 检验来隔离不同的试验台。

#### 第四章- 第 C 节 变差研究

所有描述的研究，是在研究期间对测量系统（包括环境的后果）本质的计算。由于测量系统是被用于作出对产品、过程或服务的未来判定，所以就需要对测量系统作出分析的结论。从计算到分析结果的转化需要专业知识和专家，以能：

- 确保某研究的设计和已实施已考虑到所有预期的测量变差来源。
- 根据预期的用途、环境、控制，保养等分析结果（数据）。

---

#### V1: 标准 GRR 研究

这些研究已被包含在本参考手册中。  
这些研究包括图形分析和数值分析。

- V1a——极差法（R&R）
- V1b——极差法（R&R 和零件内部）
- V1c——ANOVA 法（R&R）
- V1d——修改的 ANOVA/极差法

---

#### V2:p≥2 台仪器的多个读数

这里允许多台仪器的比较。

##### 应用：

a) 零件不会由测量过程而被改变的测量系统，即测量系统是非破坏性的，并且可使用具有如下特性的零件（样本）：

- 1) 静态属性，或
- 2) 稳定的动态（变化）特性

##### 假设：

- 已知特性（属性）的定性期限，并延续超过预期的研究期间，即在整個预期的使用和/或储存期间，被测特性不会改变；

- 零件（样本）涵盖了该特性（属性）的过程变差的预期范围；

采用 Grubb<sup>70</sup>（或 Thompson）<sup>71</sup> 估计的分析：

- 过程变差性；
- 仪器变差性=重复性；
- 置信度范围的计算是可获得的。

### V3：分割的样本（m=2）

#### 应用：

个别的零件（样本）部分的测量不可被重复，因此可用破坏性和不可重复的测量系统进行研究，并且可被使用在动态我的测量系统分析。

#### 假设：

- 已知特性（属性）的定性期限，并延续超过预期的研究期间；即在整個预期的使用和/或储存期间被测特性不会改变；
- 零件（样本）涵盖了该特性过程变差的预期范围；
- 这些样本被分成 m 部分。m=2 时常称为试验一再试验研究。

#### 采用回归技术分析：

- 用误差估术语计重复性： $\sigma_{\text{重复性估计}} = \sigma_e^2$ ；
- 线性（通过估计线和 45° 线进行比较）

V3a—V3 用于连续的配对零件

本研究同 V3 一样，但使用连续连续零件的配对，而不是分割样本。本研究应用在当不破坏被测量的属性，该零件就不能被划分的情况。

这是一个上边界分析： $\sigma_{\text{重复性估计}} \leq \sigma_e + \sigma_{\text{样本之间}}$

### V4—分割样本（一般）

#### 应用：

个别的零件（样本）部分的测量不可被重复，因此可用破坏性和不可重复的测量系统进行研究，并且可被使用在动态特性的测量系统分析。

<sup>70</sup> 见参考清单，NO.15。

<sup>71</sup> 见参考清单，NO.38。

#### 假设:

- 特性（性质）的保存寿命已知并延续超过预期的研究时期，即在预期的使用和/或储存期间被测特性不改变；
- 零件（样本）涵盖了该特性（属性）的过程变差的预期范围；
- 把样本分割成  $m$  部分。这里  $m$  为 2 或 3 的倍数；  $m \geq 2$ （如  $m=3, 4, 6, 9\cdots$ ）

#### 采用以上方法进行分析:

包括图形法的标准 GRR 技术

- ANOVA——随机化的分组设计（双向 ANOVA）。

#### V4a—V4 用于从不同批次中连续（同质的）配成对的零件

本研究同 V4 研究一样，但使用连续零件而不是分割样本。本研究应用在于当不破坏被测量的属性，该零件就不能被划分的情况。

这是一个上边界研究： $\sigma_{\text{重复性估计}} \leq \sigma_e + \sigma_{\text{零件}} + \sigma_{\text{批次}}$

下述研究件假定零件（样本）特性（属性）是动态的：

---

#### V5: 与 V1 相同，使用稳定的零件

本研究中使用基于工程知识和专门技术为基础的稳定过程的零件；例如，磨合过的发动机对新发动机的比较。

---

#### V6—时间序列分析

##### 假设:

- 在规定的时间内重复的读值；
- 已知特性（属性）的定性期限，并延续超过预期的研究期间；即在预期的使用和/或储存期间，被测特性不会改变；
- 零件（样本）涵盖了该特性（属性）的过程变差的预期范围；

通过确定了每个样本的退化模型来分析:

- $\sum_{\text{重复性估计}} = \sigma_e$  ;
- 退化的一致性（如果  $n \geq 2$ ）;



---

## V7: 线性分析

假设:

- 在规定的时间内重复的读值;
- 已知(已有文件说明的)在规定的时间内, 这测量系统中的退化具有线性回应。
- 已知特性(属性)的定性期限, 并延续超过预期的研究期间; 即在预期的使用和/或储存期间, 被测特性不会改变;
- 零件(样本)涵盖了该特性(属性)的过程变差的预期范围;

通过线性回归来分析:

- $\sum \text{重复性估计} = \sigma_e$  ;
- 退化的一致性 (如果  $n \geq 2$ );

V7a—V7 用于同质性的样本

通过线性回归来分析:

这是一个上边界分析:  $\sigma_{\text{重复性估计}} \leq \sigma_e + \sigma_{\text{零件之间}}$

---

## V8: 特性(性质)随时间的衰变

可以修改 V6 和 V7, 以确定这退化是否基于于时间(即定性期限)或活动

---

## V9: V2 用于同时存在的多个读值, 且 $P \geq 3$ 台仪器

分析同 V2[参见 Lipson 和 Sheth, 13.2 节]。

## 第 D 节

### 识别过大的零件内部变差的影响

了解一测量系统的变差来源对于所有测量的应用是很重要的，特别是当有一个显著的零件内变差时。零件内部变差，如锥度或失去真圆度，会造成测量系统的评价提供误导的结果；这是因为未考虑零件内变差对重复性、再现性或两者的估计。也就是说，零件内部变差可能成为测量系统变差的一个重要因素。就目前工作来说，了解产品中呈现的零件内部变差，将比了解测量系统的适用性有更大的意义。

可能遇到的零件内变差的例子有：真圆度（失去真圆度）、中心度、锥度、平面度、表面粗糙度、圆柱度等。<sup>72</sup> 在同一零件中同时存在一个以上的这些特性（组合的误差）是可能的。每个特性的强度和它们的相互依赖关系可能包含在数据中，并且影响对测量系统的理解。在测量系统分析过程中最重要的，是识别和分析这些额外的变差来源，以理解测量系统的真实变差。在这种情况下，用一个 D.O.E.、ANOVA 或其它更复杂的统计技术进行分析可能是必要的。不论采用哪种技术，均应该征得顾客供方质量代表的核准。

另外，产品设计工程师所做的决定，会对如何进行零件测量、如何设计夹具产生很大的影响，而且其结果会影响最终的测量误差。例如：塑胶产品的分模线是一很关键的特性（两个模穴的模具通常有多余的塑胶材料注入，因此分模线是一个不好控制的表面）。这些因素要在设计 FMEA 过程中予以考虑。

一旦零件内变差要素，就有可能在测量系统内控制这些因素（例如，重新设计量具或使用不同的夹紧方法/设备），以使得未来的数据不致被混淆。

---

<sup>72</sup> 许多这样的特性是在设计中通过几何尺寸和公差（GD&T）来控制的。GD&T 以一种功能的方式提供一种可操作地定义的方法来检查零件。一般来说，功能检查是一种记数型数据检查。当需要计量型数据时，会产生使用为功能检查而设计的量具以得出计量型数据。这可能通常通过使用功能量个作为一个支持夹具用于 CMM 研究。然而这样做的关键是夹具要牢牢地并可重复地在相同的位置夹紧零件（如果不是这样，MSA 研究结果就会产生这种误差。

## 第五章- 第二节 平均值和极差法—额外的处理

### 引言

关于测量系统评估的平均值和极差法，有一附加考虑值得一提。一控制图的范例经允许摘自于 Wheeler 和 Lyday 所著的“测量过程评估“(见参考文献)。

如其它的优秀测量系统分析设计一样，这种图形方法的主要目的是：确定某测量过程是否足以测量生产过程的变差和/或评价产品性能：

- 所有量具在做相同的工作吗？
- 所有评价人在做相同的工作吗？
- 顾及过程变差，测量变差是否可以接受？
- 从测量过程获得的数据有多好？或这些数据能划分成多少非重迭组成或类别？

### 程序步骤

- 1) 应该小心地按第二章第 C 节“测量系统研究的准备”进行。
- 2) 让每位评价人测量将被研究其特性的每个样本。在控制图最上面的的数据行中记录首次检查结果（见图 33）。
- 3) 重复测量并在控制图上的第二数据栏中记录数据。（注：当进行第二次测量时，不允许评价人看到他们的原始读值。）现在由每位评价人对同一零件应有两次测量数据。
- 4) 通过计算各子组的均值（ $\bar{X}$ ）和极差（ $R$ ）来分析数据。
- 5) 在极差图上描绘出极差值并计算平均极差（ $\bar{R}$ ）（包括所有评价人的所有子组极差（ $R$ ））。在图上画出该极差平均值。有  $n=2$  时的  $D_4$  常数以计算这极差图的控制限值。画出这限值并确定所有数值是否都受控。

- √ 如果所有极差都受控，则所有评价人在做同样的工作。
- √ 如果一位评价人失控，则他的方法与其他人不同。
- √ 如果所有评价人都有一些失控的极差值，则该测量系统对评价人的技术是敏感的，且需要改进以获得有用数据。

- 6) 下一步，在平均值图上描绘所有评价人的各子组均值（ $\bar{X}$ ）（见图 33）。该平均值代表了零件变差和测量变差。

计算总平均值（ $\bar{X}$ ）（包括所有评价人的所有子组平均值（ $\bar{X}$ ））。在图上画出该总均值（ $\bar{X}$ ）。

现在，有  $n=2$  时的  $A_2$  因子和极差图的平均极差（ $\bar{R}$ ）计算该图的控制限并在均值图上画出这

些限值。注意，本项研究中的极差图只包含测量变差。因此，平均值的控制限之间的面积代表在这系统的测量变差量。

如果所有平均值落在控制限值内，则这测量变差蒙蔽了该过程变差。换句话说，这测量过程比生产过程有更大的变差，监控和控制该生产过程没有价值。

如果少于一半的均值落在控制限外，则这测量系统不适用于过程控制。

相反的，如果大多数平均值落在控制限外，则表明来自生产过程的信号大于测量变差。这测量系统能提供用于控制该过程的有用数据。

### 计算表举例

问题——“这测量系统所收集的数据好到什么程度？”这个问题，可以通过完成图 34 的工作表示例后得到回答。这计算表所需的所有数据都可在上述的平均值和极差图中找到。

以下是完成工作表示例（图 34）的步骤：

- 1) 识别将被评价的测量和特性，谁是实施这研究的执行者和研究的日期。
- 2) 直接从控制图上得到子组平均极差 ( $R$ )。
- 3) 重复次数 ( $r$ ) 是每个评价人检查同一零件的次数。
- 4) 如图 34 所示，采用与  $r$  对应的  $d_2$  值计算估计的重复性标准差 ( $\sigma_E$ )
- 5) 在提供的空白处位置填入评价人数 ( $n_A$ )。
- 6) 在提供的位置填入样本数量 ( $n$ )。
- 7) 平均每位评价人所获得的所有样本，计算每位评价人的均值，并把这些均值填入提供给各评价人 (A, B, C) 的空白处。
- 8) 检查每个评价人 (A, B, C) 的平均值，由最高值减少去最低值来确定评价人平均值的极差，并填入 ( $R_A$ ) 的空白处。
- 9) 采用  $n_A$  值对应的  $d_2^*$  值计算估计的评价人标准偏差 ( $\sigma_A$ )。
- 10) 通过平均每个样本的所有评价人所获得的数值，计算样本平均值。(例如，评价人 A 的样本 1 平均值+评价人 B 样本 1 平均值+最后一位评价人样本 1 的平均值，其总和除以评价人人数量)。这是样本真值的最佳估计。在图 34b 中所提供的空白处 (1, 2, 3.....9, 10) 填入每个样本的平均值。
- 11) 观察样本均值 (1, 2, 3.....9, 10)，并通过最高值减去最低值，计算样本均值的极差 ( $R_p$ )，把该值填入提供的空白处。
- 12) 采用  $n$  值 ( $n$ =样本数量) 对应的  $d_2^*$  值，计算估计的样本间的标准偏差 ( $\sigma_p$ )。
- 13) 通过样本标准差除以测量标准偏差，以计算“信噪比”(SN)，并填入提供的空白处。  
$$SN = \sigma_p / \sigma_{GRR}$$
- 14) 确定区隔产品的区别分类数，这能由这些测量来区隔。将 SN 乘以 1.41 并填入提供的空白处 (图 34b)。

只需考虑它的整数部分，因为这定义了区别的分类。(去掉所有的小数部分。)(见图 34b。)

如果分类数量小于 2，则测量系统对于控制该过程毫无价值；它全是干扰，一个零件不能被指为与另一个零件不同。

如果分级数等于 2，则意味着数据只能分成高、低两组，这只相当于计数型数据。

如果分级数等于 3，数据能分成高、中、低三组。这是一个稍好一些的测量系统。包含 4 个或更多分级的系统会比前面三个例子好得多。

图 33a: 测量评价控制图 ( $\bar{X}$ &R) <sup>-1</sup>

图 33b: 测量评价控制图 ( $\bar{X}$ &R) -2

评价的测量 接头 完成人 R.W.L 日期 MM-DD-YY

重复误差: 子组极差均值  $\bar{R} = 7.6$

重复次数=子组大小  $= r = 2$

估计的重复性标准偏差  $\frac{\bar{R}}{d_2} = \hat{\sigma}_e = 6.738$

$r$	$d_2$
2	1.126
3	1.693
4	2.059
5	2.326
6	2.534
7	2.704
8	2.847
9	2.970
10	3.078

评价人影响的计算:

评价人均值如下:

$n_A$	$d_2^*$
2	1.410
3	1.906
4	2.237
5	2.477
6	2.669
7	2.827
8	2.961
9	3.076
10	3.178

评价人数  $= n_A = 3$

样品数  $= n = 5$

$108.1 - 102.8 = R_A$

评价人均值的极差  $= R_A = 5.3$

$\frac{5.3}{1.906} = \hat{\sigma}_0$        $\frac{\bar{R}}{d_2^*} = \hat{\sigma}_0 = 2.781$

评价人	均值
A	103.1
B- 低	102.8
C- 高	108.1
D	—
E	—
F	—

图 34a: 评价测量过程的控制图方法的计算 (第 1 部分, 共 2 部分)

测量误差标准偏差的计算：

估计测量误差标准偏差是：

$$\sigma_m = \sqrt{\sigma_e^2 + \sigma_0^2 + \dots}$$

**图 34b：评价测量过程的控制图法的计算（第 2 页，共 2 页）**



## 第 F 节 量具性能曲线<sup>73</sup>

建立一量具性能曲线(GPC)的目的是确定接受还是拒收某零件一些参考值的概率。理论上,一个没有误差的测量,其 GPC 如图 35 所示。但这是对测量系统的理想而不是正常会发生的情形。

一旦确定了误差量,当采用某系统,就有可能要计算一些其接收某零件参考值的概率。

为完成这工作,首先假设测量系统误差主要由差的重复性、再现性和偏倚组成,再进一步假设重复性和再现性为正态分布,另外有一些变异为  $\sigma^2$ 。量具的误差是平均值  $X_T$ (即参考值)加上偏倚和一些方差  $\sigma^2$  的正态分布。换句话说:

由量具得到的实际值= $N(X_T + b, \sigma^2)$   
接受某零件参考值的概率表示成如下关系式:

$$P_a = \int_{LL}^{UL} N(X_T + b, \sigma^2) dx$$

采用标准正态分布表:

$$P_a = \int_{LL}^{UL} N(X_T + b, \sigma^2) dx$$

式中:

---

<sup>73</sup> 经允许,节选自 J.McCaslin 和 G.Grusks 的“计算型量个系统分析”,ASQC,1976。

## 范例:

当参考扭矩值为 0.5Nm, 0.7Nm, 0.9Nm 时, 确定接受一零件的概率。

采用前述所进行研究中的数据:

$$\text{规范上限}=\text{UL}=1.0\text{Nm}$$

$$\text{规范下限}=\text{LL}=0.6\text{Nm}$$

$$\text{偏倚}=\text{b}=0.05\text{Nm}$$

$$\sigma_{\text{GRR}}=0.05\text{Nm}$$

将上面数据代入前一页的公式:

$$P_a =$$

$$P_a =$$

$$P_a =$$

$$P_a = 1.0 - 0.84 \\ = 0.16$$

当该零件有一参考值为 0.5Nm 时, 大约有 84% 的机会将被拒收。

## 量具性能曲线范例

对于  $X_T=0.7\text{Nm}$

$$P_a =$$

$$P_a =$$

$$P_a = 0.999$$

如果该零件的参考值为 0.7Nm, 则它被拒收的机会少于 0.1%。

当  $X_T=0.9\text{Nm}$

$$P_a =$$

$$P_a =$$

$$P_a = 0.84$$

如果该零件的参考值为 0.9Nm，则它被拒收的机会少于 16%。

如果对于所有  $X_T$  值计算了接受概率并画成图形，然后就得到了量具性能曲线，如图 36 所示。

在正态概率纸上更容易绘制这相同的曲线，如图 37。一旦绘出图形，这 GPC 曲线能得到对任何零件尺寸的一个零件接收的概率。

而且，一旦 GPC 被建立，它可被用于计算重复性和再现性误差以及偏倚误差。<sup>74</sup>

这 5.15GRR 范围可以经由找出概率  $P_a = 0.995$  对应的  $X_T$  值，和  $P_a = 0.005$  对应的  $X_T$  值，而确定两个限值。GRR 就是这两个  $X_T$  值的差值，如图 37 图形所示。

偏倚的估计值是经由找出对应通过找出对应于  $P_a = 0.5$  的上限值或下限值的  $X_T$ ，并计算：

$$B = X_T - LL \text{ 或 } BB = X_T - UL$$

取决于选择哪个  $X_T$  限值。<sup>75</sup>

---

<sup>74</sup> 见“计数型量个研究”第三章第 C 节

<sup>75</sup> 这假定测量系统在其工作量程范围内是线性的。

偏倚=0.00, GRR 极差=0.00

图 35: 量具性能曲线----无误差

被测零件的基准值

偏倚=0.05 GRR 极差=0.24

图 36:量具性能曲线----示例

图 37:绘制在正态概率纸上的量具性能曲线

## 第 G 节

### 通过多次读数减少变差

如果目前的测量系统变差是不可接受的(大于 30%),有一种方法可用来将变差减少到可接受水准,直到已对测量系统进行了适当的改进。通过对被评价零件特性采取多次统计上独立(非相关)的测量读值,确定这些测量的平均值,并以这结果的数值代替各次测量,这样可能减少了不可接受的变差。这种方法当然会费时,但是一种替代的方法,直至已对量具系统进行改进(如重新设计或购买一个新量具)。这种替代方法的程序如下:

- 1) 确定为了满足变差的可接受水准,必要的多次读数的数量。
- 2) 遵循在这本手册前面所讨论的量具研究程序。

范例:

在 XTZ 公司的范例中,公差 R&R 变差百分比为 25.5%, 有一个  $6\sigma$  宽度为 0.24。顾客要缩减该数值至最少 15%,其将有一个等同的  $6\sigma$  宽度为 0.14。<sup>76</sup>

为了确定满足希望的 15%准则,确定所需的多次读值数量,首先必须明白,个别和平均测量值的分布具有相同平均值,其次,平均值分布的变异等于个别分布的变差除以样本大小。了解了以下所示的这种关系,就能确定所需多次读值数量。

$$(\sigma_{\bar{x}})^2 = \frac{(\sigma)^2}{n}$$

同时,该等式也可以表示为:

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

以及

$$0.14 = \frac{0.24}{\sqrt{n}}$$

所以

$$\sqrt{n} = 1.714$$

于是

$$n = 3(\text{四舍五入至最接近的整数})$$

因此,该零件特性的 3 次读数将使整个测量系统的变差减少到约为 0.14,%GRR 减少到 15%。

这种方法应该被视为一个临时步骤,直至已对测量系统进行其它改进。该方法应该只有在顾客同意的情况下被使用。

---

<sup>76</sup> 见 IV 页注。

## 第 H 节 聚集标准差法计算 GRR<sup>77</sup>

G 测量系统的分析通常假设所有的评价人以随机的方式获得了所有零件/样本的可重复的数据。但这种假设不一定总是可能的。如果某测量系统分析了包括多个场所。逻辑上这很难做到随机的抽样。同样对一些试验，特别是化学和金相分析（实验室之间和实验室内部研究），可能需要采用不同样本的混和取样，这些样本不是一同质过程的零件，也可能不是在同一时间获得的。

这些情况可以使用一群体的 DOE 来处理这一问题。一种替代的方法是聚集的标准差法 GRR 研究，这是遵循在 ASTM E691 所讨论过的方法。

这种方法将每个零件看成一个独立的材料，然后如 E691 计算重复性和再现性标准差。这样，将产生多个单独的重复性和再现性值。由于被研究的零件的标识是不可或缺的，假设这些单独的估计值也被有效地标识。当然，它们从来不会相同，但它们的平均值将对重复性的真实水准，提出一个很好的估计值。

如果这些方法用于评价一组实验室，有一个顾虑是：“再现性”的本质到底是什么？如果大多数时间均大于零（也就是大多数的材料均大于 0），这就意味着应该解释为，有一个存在于操作者之间差异，例：对于内部实验室计划来说，实验室之间存在真正的差异。

### 连续的应用

尽管这 E691 方法通常是用于完整样本的情况，但也可把它用在连续样本的方法；对当所有样本不能在同一时间获得的情况下是很有用的。另外，它还可用于校准过程的一部分，以保持测量系统变差的信息。

下面的研究假定该研究可以用于连续的情况。

### 进行研究

应严格按照第二章第 C 节所述的“测量系统研究的准备”来进行。

---

<sup>77</sup> 本节的部分内容包括美国试验和材料协会的 Neil Ullman 提供的“一致性统计学”的所有内容（国际的 ASTM）。

**接续自第二章第 C 节的第 6 步骤:**

- 7) 让每个  $m \geq 2$  的评价人评估零件  $r \geq 3$  次读数。将数据记录在数据收集表 (参见范例表格) 的适当行中。(注意, 在进行多次读值时, 不要让评价人看到他们以前测量的读值。)
- 8) 计算每个评价人每个新零件的均值 ( $\bar{X}$ ) 和标准偏差 ( $S$ )。
- 9) 将标准偏差图数值描绘在标准差图表上, 并计算标准偏差的平均值 ( $\bar{S}$ ) (包括对所有评价人的所有子组的标准差)。在图上画出这个标准偏差均值, 用对应测量次数  $r$  的  $D_4$  常数计算标准偏差图表的上控限制, 画出该控制限并判断是否所有值都受控 (见图 38)。
- 10) 将所有评价人的每个子组的平均值 ( $\bar{X}$ ) 画在平均值图表上 (见图 38)。平均值代表过程变差和测量变差两者。
- 11) 计算总平均值 ( $\bar{X}$ ) (包括所有评价人的所有子组均值 ( $\bar{X}$ ))。在平均值图表上画出该总平均值。
- 12) 用对应  $r$  的  $A_2$  常数, 以及从标准差图上的标准差均值 ( $\bar{S}$ ) 计算该图的控制限, 在均值图上画出这些限值。
- 13) 应用控制图和再平均值和极差图 (见第三章) 中所讨论的其它图形方法来分析数据。
- 14) 通过聚集了评价人的结果, 评价每个零件评价测量系统的参数。

$$S_{\bar{x}_g} = \sqrt{\frac{\sum_{i=1}^m S_{xi}^2}{m}}$$

$$\text{重复性}_g = S_{Ag} = \sqrt{\frac{\sum_{i=1}^m S_{Si}^2}{m}}$$

$$\text{再现性}_g = S_{Ag} = \sqrt{S_{Eg}^2 - \frac{S_{Eg}^2}{m}}$$

$$GRR_g = S_{GRRg} = \sqrt{S_{Eg}^2 + S_{Ag}^2}$$

E691 的按照在 MSA 中的惯例, 将再现性定义为评价人的变差, 以及 MSA 的 GRR 称为再现性。在这种情况下:

$$S_{\text{评价人}} = \sqrt{S_x^2 - \frac{S_r^2}{3}}$$

$$S_R = \sqrt{S_r^2 + S_{\text{评价人}}^2}$$

其中  $S_r = S_E = \text{重复性}$ ,  $S_r = GRR = \text{ASTM 再现性}$



15) 通过聚集了零件结果，评价整体测量系统的参数。

$$\text{重复性} = S_E = \sqrt{\frac{\sum_{i=1}^g S_{Si}^2}{g}}$$

$$\text{再现性} = S_A = \sqrt{\frac{\sum_{i=1}^g S_{Ai}^2}{g}}$$

$$\text{GRR} = S_{GRR} = \sqrt{\frac{\sum_{i=1}^g S_{GRRi}^2}{g}}$$

当计算总变差百分比时，过程的历史标准差应该被使用。

如果这零件涵盖了一个广泛多变的过程，例如：不同的金相和化学样本，评价总变差的百分比时，应该以指定样本的过程变差为基础，而不是所有样本的总变差为基础。

当评价人是隶属于不同的场所（如试验室）时，在解释测量系统的参数时，应该特别注意。

重复性将包含设备间的变差以及设备内部的变差。它可以通过计算和比较每个场所的重复性来进行评价。

再现性将包括不同场所之间和评价间的变差。在本研究中，不能将这两个要素分开。

### 回应的 X/S 图

图 38a: 聚焦的标准差研究图示分析



## 一致性统计

ASTM 和 ISO 方法<sup>78</sup> 建议计算两个“一致性”统计量，即 h 和 k。h 值用下述公式计算：

$$h = \frac{X_{\text{评价人}} - X_{\text{零件}}}{S_x}$$

对于评价人 A 和零件 1 来说，平均值（上式  $X_{\text{评价人}}$ ）为 0.447，零件的均值（上式  $X_{\text{零件}}$ ）为 0.169。评价人间的标准偏差（上式  $S_x$ ）为 0.262。则

$$h = \frac{0.447 - 0.167}{0.262} = \frac{0.278}{0.262} = 1.06$$

k 值是每个评价人测量每个零件的标准差与重复性标准偏差的比值。在本例中(评价人 A 和零件 1),它是：

$$k = \frac{\text{标准差(评价人 A, 零件 1)}}{\text{重复性}} = \frac{0.178}{0.132} = 1.35$$

这些计算的原因之一是允许对非常不同的材料进行比较。

尽管在本例中没有收集差别较大的材料，这些材料有不同的水准，而且它们的标准差可能相差非常大。但仍可应用 E691 中的 h 和 k 的计算，以比较重复性标准差和评价人的回应数值。下表列出各评价人的 h 和 k 值。

评价人	h										平均 h	“z”
	1	2	3	4	5	6	7	8	9	10		
A	1.06	0.87	0.82	0.86	0.88	0.32	0.65	0.80	0.69	1.05	0.80	2.53
B	-0.14	0.22	0.29	0.24	0.21	0.80	0.50	0.32	0.46	-0.11	0.28	0.88
C	-0.93	-1.09	-1.11	-1.10	-1.09	-1.12	-1.15	-1.12	-1.15	-0.94	-1.08	-3.41

评价人	k										中值 k	“z”
	1	2	3	4	5	6	7	8	9	10		
A	1.35	0.25	0.49	0.28	0.23	0.65	0.59	0.35	0.77	0.27	0.42	-3.20
B	0.77	1.50	1.15	1.70	1.50	1.20	1.40	1.68	1.07	0.45	1.30	3.14
C	0.76	0.82	1.20	0.14	0.84	1.07	0.83	0.24	1.13	1.65	0.84	-0.17

上表的最后两列是均值和“z”值，用来观察评价人是否为显著不同。在他们对零件大小的读值中，h 值表明评价人 A 明显地大，评价人 C 明显的小。正是这种明显差异产生 GRR 的标准差。

<sup>78</sup> 见 ISO5725。

重复性标准差也可以通过  $k$  值的方法评价。为此，计算  $k$  的中间值，然后得出一个估计的“ $z$  分”。本研究中，期望的中间值为 0.861，其标准偏差约为 0.439。评价人 A 的中值  $k$  约为在期望的水平之下的-3.2 标准偏差，而评价人 B 显著地大，因此我们在仅有三个操作者的表现中就能看到很大的差别。

$h$  图（图 38b）和  $k$  图（图 38c）也有助于说明这些差别。评价人 C 比其它人有更小的结果。同样的， $k$  值表明评价人 A 的重复性变差如何小。这些是在检查由评价人进行的关于测量方法的表现时必然检查的事项。

**图 38b:  $h$  值的点图**

**图 38c:  $K$  值点图**

# 附录

## 附录 A 方差分析的概念

可依据表 18 中的公式进行对 GRR 数值的分析，这就是被称为“方差分析”(ANOVA)表。ANOVA 表包括六列。

- 来源列为变差的原因。
- DF 列为与变差来源有关的自由度
- SS 或平方的总和列为来源平均值附近的偏差。
- MS 或平均平方列是平方的总和除以自由度。
- EMS 或预期的平均平方列确定了每个 MS 的变数元素的线性组合。ANOVA 表将总变差来源分解成为四个要素：零件、评价人、评价人和零件的交互作用，以及量具/设备重复性产生的重复误差。
- F 比率列只用于计算的目的仅是为了说明一种 MSA ANOVA 中的交互作用，这是由交互作用的均方除以均方误差来确定的。

对每个误差来源的变数元素估计值在表 16 得出。<sup>79</sup>

---

量具 (EV)	变数估计值
相互作用 (INT)	$\tau^2 = MS_e^{80}$
	$\gamma^2 = \frac{MS_{AP} - MS_e}{r}$
评价人 (AV)	$\omega^2 = \frac{MS_A - MS_{AP}}{nr}$
零件 (PV)	$\sigma^2 = \frac{MS_P - MS_{AP}}{kr}$

---

表 16: 变数元素的估计值

因为每个平均平方值是一个受抽样变差影响的一个样本数量的值，并且这计算涉及到平均平方的差异，所以可能出现负的变差要素估计值。因为“基准”变数要素等于或接近于 0，或者有一个小的样本大小，所以这是一点小问题。为了分析，负的变数要素被设定为 0。

---

<sup>79</sup> 本表中，假定所有方差分量具有随机效果。

<sup>80</sup> 在本 ANOVA 对测量系统分析的应用中，ANOVA 误差术语等同于 MSA 设备变差， $MS_e$ 。

标准偏差比变差易于解释，因为它与原始观测使用相同的测量单位。实际上，基本分布宽度的测量被设定为 5.15 倍的标准偏差<sup>81</sup>。表 17 列出了被称为设备变差 (EV) 的重复性度量和被称为评价人变差 (AV) 的再现性度量，两者利用了  $5.15\sigma$  分布宽度的公式。如果零件和评价人的相互作用是显著的，则存在一个非附加的模型，于是可以提出它的变数要素的估计值。表 17 中的 GRR 是总的测量系统变差。

---

$EV = 5.15\sqrt{MS_e}$	设备变差=重复性
$AV = 5.15\sqrt{\frac{MS_A - MS_{AP}}{nr}}$	评价人变差=再现性
$I_{AP} = 5.15\sqrt{\frac{MS_{AP} - MS_e}{r}}$	评价人与零件的交互作用
$GRR = \sqrt{(EV)^2 + (AV)^2 + (I_{AP})^2}$	量具 R&R
$PV = 5.15\sqrt{\frac{MS_P - MS_{AP}}{kr}}$	零件变差

---

**表 17:  $5.15\sigma$  分布宽度**

在附加的模型中，交互作用不显著，且各来源的变数要素被确定如下：首先，将量具误差的平方和（表 18 中  $SS_e$ ）加上评价人与零件交互作用的平方和（表 18 中  $SS_{\text{评价人}}$ ）这等于具有  $(nkr - n - k + 1)$ <sup>82</sup> 自由度的聚集平方总和 ( $SS_{\text{聚集}}$ )。然后用 ( $SS_{\text{聚集}}$  除以  $(nkr - n - k + 1)$ ) 来计算  $MS_{\text{聚集}}$ 。 $5.15\sigma$  分布宽度限值将为：

$$EV = 5.15\sqrt{MS_{\text{聚集}}}$$

$$AV = 5.15\sqrt{\frac{MS_A - MS_{\text{聚集}}}{nr}}$$

$$GRR = \sqrt{(EV)^2 + (AV)^2}$$

$$PV = 5.15\sqrt{\frac{MS_P - MS_{\text{合并}}}{kr}}$$

---

<sup>81</sup> 这是 99% 范围。见 IV 页注。

<sup>82</sup> 其中 n=零件数，k=评价人数，r=试验次数。

为了确定交互作用是否显著，计算零件对评价人的相互作用的 F 统计（见表 18）。将这 F 与取自 ANOVA(表 18)中以自由度为分母、F 分布为分子的一个上端百分比点相比较。

为了减小错误地作出没有交互作用的结论的风险，选择了一个较高的显著性水平。一旦确定 GRR，则可以计算同过程性能有关的 %GRR。

$$SS_{\text{零件}} = \sum_{i=1}^n \left( \frac{X_{i\cdot}^2}{kr} - \frac{X_{i\cdot}^2}{nkr} \right)$$

$$SS_{\text{评价人}} = \sum_{j=1}^n \left( \frac{X_{\cdot j}^2}{nr} - \frac{X_{\cdot j}^2}{nkr} \right)$$

$$TSS = \sum_{i=1}^n \sum_{j=1}^k \sum_{m=1}^r \left( X_{ijm}^2 - \frac{X_{i\cdot}^2}{nkr} \right)$$

$$SS_{\text{零件对评价人}} = \sum_{i=1}^n \sum_{j=1}^k \frac{X_{ij\cdot}^2}{nr} - \sum_{i=1}^n \frac{X_{i\cdot}^2}{kr} - \sum_{j=1}^k \frac{X_{\cdot j}^2}{nr} + \frac{X_{\cdot\cdot}^2}{nkr}$$

$$SS_e = TSS - [SS_A + SS_P + SS_{AP}]$$

变差源	DF	SS	MS	F	EMS
评价人	k - 1	SS <sub>A</sub>	MS <sub>A</sub> = $\frac{SS_A}{(k-1)}$		$\tau^2 + r\gamma^2 + nr\omega^2$
零件	n - 1	SS <sub>P</sub>	MS <sub>P</sub> = $\frac{SS_P}{(n-1)}$		$\tau^2 + r\gamma^2 + kr\sigma^2$
零件的评价人	(n-1)(k-1)	SS <sub>AP</sub>	MS <sub>AP</sub> = $\frac{SS_{AP}}{(n-1)(k-1)}$	$\frac{MS_{AP}}{MS_e}$	$\tau^2 + r\gamma^2$
设备	nr(n-1)	SS <sub>E</sub>	MS <sub>E</sub> = $\frac{SS_e}{nk(r-1)}$		$\tau^2$
总变差	nkr-1	TSS			

评价人 ~N(0, ω<sup>2</sup>)  
零件 ~N(0, σ<sup>2</sup>)  
X 零件的评价人 ~N(0, γ<sup>2</sup>)  
设备 ~N(0, τ<sup>2</sup>)

表 18 变差数分析 (ANOVA)

表 19a 和表 19b 是用 ANOVA 法计算在图 24 中的举例数据。

变差来源	DF	SS	MS	F	EMS
评价人	2	3.1673	1.58363	34.44*	$\tau^2+3\gamma^2+30\omega^2$
零件	9	88.3619	9.81799	213.52*	$\tau^2+3\gamma^2+9\sigma^2$
零件的评价人	18	0.3590	0.01994	0.434	$\tau^2+3\gamma^2$
设备	60	2.7589	0.04598		$\tau^2$
总变差	89	94.6471			

\*显著性是在  $\alpha=0.05$  水准

表 19a: ANOVA 结果的列表

变差的估计	5.15( $\sigma$ )			%总变差	%贡献度
	标准差( $\sigma$ )				
$\tau^2=0.039973$ (设备)	0.199933	$EV=1.029656$		18.4	3.4
$\omega^2=0.051455$ (评价人)	0.226838	$AV=1.168213$		20.9	4.4
$\gamma^2=0$ (相互作用)		$INT=0$		0	0
$GRR=0.09143$ $\tau^2+\gamma^2+\omega^2$	0.302373	$GRR=1.557213$		27.9	7.8
$\sigma^2=1.086447$ (零件)	1.042327	$PV=5.367987$		96.0	92.2
总变差	1.085	$TV=5.589293$		100.0	

表 19b: ANOVA 结果的列表  
(变差的估计是基于无交互作用的模型)

$$ndc = 1.41(PV/GRR)=1.41(5.37/1.56)=4.85 \cong 4$$

$$\text{总变差}(TV) = \sqrt{GRR^2 + PV^2}$$

$$\% \text{总变差} = 100 \left( \frac{5.15\sigma_{\text{要素}}^2}{5.15\sigma_{\text{全部}}^2} \right)$$

$$\% \text{贡献(对总变差)} = 100 \left( \frac{5.15\sigma_{\text{要素}}^2}{5.15\sigma_{\text{全部}}^2} \right)$$



## 附录 B

### GRR 对能力指数 $C_p$ 的影响

公式:

$$\sigma_o^2 = \sigma_A^2 + \sigma_M^2 \quad (1)$$

式中: O---观测的过程变差  
A---实际的过程变差  
M---测量系统变差

$$CP_x = \frac{|U - L|}{6\sigma_x} \quad (2)$$

式中: U, L 是规范的上下限值  
x=O 或 A, 其意义见 (1)

$$GRR_p \% = GRR_p * 100\% \quad (3)$$

基于过程变差:

$$GRR_p = \frac{k\sigma_M}{6\sigma_o} \quad (4)$$

注意:  $GRR_p \leq 1$ , 因为按照公式 (1),  $\sigma_o^2 \geq \sigma_M^2$

基于公差范围:

$$GRR_p = \frac{k\sigma_M}{|U - L|} \quad (5)$$

在公式 (4) 和 (5) 中, k 通常取 5.15。然而, 为了没有一般性的损失, 这个分析中 k 将取 6, 以便简化计算。

分析:

$$\begin{aligned} CP_o &= CP_A * \frac{\sigma_A}{\sigma_o} \\ &= CP_A * \frac{\sqrt{\sigma_A^2 - \sigma_M^2}}{\sigma_o} \quad \text{用公式 (1)} \end{aligned}$$

用基于过程变差的 GRR

$$\begin{aligned} CP_o &= CP_A * \frac{\sigma_o \sqrt{1 - GRR^2}}{\sigma_o} \quad \text{用公式 (4)} \\ &= CP_A * \sqrt{1 - GRR^2} \quad (6) \end{aligned}$$

$$\text{或 } CP_A = \frac{CP_O}{\sqrt{1-GRR^2}} \quad (6')$$

用基于公差宽度的  $GRR$

$$GRR = \frac{1}{CP_O} * \frac{\sigma_M}{\sigma_O} \quad \text{用公式 (2) 和 (5)}$$

因此:

$$CP_O = CP_A * \sqrt{1 - (CP_A * GRR)^2} \quad (7)$$

和

$$CP_A = \frac{CP_O}{\sqrt{1 - (CP_O * GRR)^2}}$$

### 图形分析:

根据公式 (6), 考虑到  $CP_A$  的  $CP_O$  直线族为:

实际的  $Cp$  值

图 39: 观测的和实际的  $Cp$  关系 (以过程为基础)

实际的	<i>GRR</i>							
	10%	20%	30%	40%	50%	60%	70%	90%
	观测 Cp 值与基于过程变差的 Cp 值							
1.3	1.29	1.27	1.24	1.19	1.13	1.04	0.93	0.57
	观测 Cp 值与基于公差的 Cp 值							
1.3	1.29	1.26	1.20	1.11	0.99	0.81	0.54	never

表 20: 观测与实际 Cp 值的对比

观测到的 Cp

图 39: 观测 Cp 与实际 Cp 关系 (以公差为基础)

附录 C--- $d_2^*$ 表

与均值极差分布相关的值

		子组容量 Size(m)																			
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
子组数	1																				
	2																				
	3																				
	4																				
	5																				
	6																				
	7																				
	8																				
	9																				
	10																				
	11																				
	12																				
	13																				
	14																				
	15																				
	16																				
	17																				
	18																				
	19																				
	20																				
$d_2^*$																					
cd																					

数据表使用：每一栏的第一行是自由度 (V)，每一栏的第二行是  $d_2^*$  值， $d_2$  是  $d_2^*$  值的无限值，额外的 V 值可以从不同的 cd 的常数得出。

说明：这个表格中的符号是按照 Acheson Duncan 的《质量控制与工业统计》，第 5 版，McGraw-Hill, 1986。

$V (R/d_2^*)^2/\sigma^2$  是当一个  $X^2$  分布于自由度 V 的近似的分布，R 是子组大小 m、子组数量 g 的平均极差。

## 附录 D 量具 R 研究

### 应用

- 只能为短期量具重复性研究提供一初步的结论。
- 可能用于量具供方交运前的检查。
- 可能用于当零件可获得量很少时的生产前准备。
- 可用于量具的开发期之中，例如：对不同的夹持置，不同的方法的快速比较。
- 没有其它的更完整和更详细的 MSA 方法一起使用时，这方法不能被用于最终的量具接收。

### 假设

- 在这时间点上，过程稳定且变差为未知。
- 线性和偏倚没有问题。
- 在此不考虑再现性，目的仅仅专注于量具的重复性。
- 用 I/MR（单值/移动极差）图来提供稳定性的最小评估。

### 用以下方法进行分析

抽取一个零件，一个操作者；将零件固定在夹具上、测量；将零件从夹具上卸下来；让同一操作者重复测量同一零件 9 次以上。

在 I/MR 图上绘制数据点，评价稳定性。如果数据表现出不稳定，实施纠正措施。<sup>83</sup>

如果稳定，用所有读值  $S$  或用  $MR/d_2^*$  来计算个别的  $\sigma$ ；乘以 6；将所得的数值除以特性的公差；再乘以 100%。评审所得到的重复性百分比同前面建立的量具接受准则相比较，或将它用为量具开发过程中的比较目的。

---

<sup>83</sup> 此处可能需要一些判断，因为单值数据的 10 个子组对建立稳定性是不够的，然而，明显的不稳定可以进行评估，并提供数值进行分析。

## 附录 E 使用误差修正术语替代 PV 计算

本手册将 PV 在此定义为  $\sqrt{TV^2 - GRR^2}$ 。由于这个零件变差的定义中包括了 EV，所以当设备变差很重要时，可能常常需要摘取这 EV 对 PV 的影响。这可通过以下公式来完成（注意：同样可用于 AV 的公式中，当 EV 的影响被摘取出来）。

$$PV = \sqrt{(R_p \times K_3)^2 - \left[ \frac{EV^2}{k \times r} \right]}$$

其中  $R_p$ =零件均值的极差,  $k$ =评价人数,  $r$ =试验次数。

用本方法计算 PV 的认知是发表于 1997 年<sup>84</sup>，这里介绍的这方法比本手册前面所介绍通用的 PV 可接受定义更具有统计上的正确性。通常，当 EV 包含 PV 时，它仅为 1% 或 2% 的程度。

<sup>84</sup> “可靠的数据是一种重要的商品”，威斯康星大学，迈迪逊的 Donald S. Ermer 和 Robin Yang E-Hok 著，出版于 ASQ 测量质量分部通讯“The Standard”中，97-1 卷，1997 年冬。

## 附录 F

### P.I.S.M.O.E.A 误差模型

与其它所有过程一样，一个测量系统受随机的和系统变差来源的影响。这些变差是由于普通原因和特殊（无次序）原因造成的。为了理解、控制和改进一个测量系统，首先要识别潜在的变差来源。尽管特定原因会取决于所处的场景，可以用一个通用的误差模型来将影响任何测量系统的变差来源进行分类。用不同的方法来说明并对这些变差来源分类，如使用简单的因果图、矩阵图或诊断树。

字头缩写 P.I.S.M.O.E.A<sup>85</sup> 是另一种用来定义测量系统变差的基本来源的模型。它不仅仅是一种模型，还可支持广泛的应用。

误差源	别名或分量	因子或参数
P 零件	生产件、样本、被测体、测试单元 (UUT)，加工品、检查标准	未知
I 仪器	量具、M&TE 单元、主量具、测量机器、试验台	比较方法
S 标准	刻度、基准、人工制品、检查标准、固有的标准、一致意见、标准基准材料 (SRM)、等级、接受准则	作为真值接受的已知值、基准值或接受准则
M 方法	岗位培训、口述、作业指导书、控制计划、检验计划、试验程序、零件程序	如何做
O 操作者	评价人、校准或试验技师、评估人、检查员	谁来做
E 环境	温度、湿度、浓度、现场清洁、照明、位置、振动、动力、电磁干扰 (EMI)、噪声、时间、空气	测量条件、噪声
A 假设	统计的、操作的、校准、常数、手册值、热稳定、弹性模数、科学定律	准则、常数或可靠测量的假设

\*实际或物理真值未知

别名误差源由于测量应用和行业的不同而不同。这些是变通的例子。然而，参数和特性影响总是相同的。

---

<sup>85</sup>P.I.S.M.O.E.A 最初由 Rock Valley 学院技术中心集成制造部主任。计量学家，ASQ 奖资深质量工程师，Gordon Skattum 提出。

所有测量系统分析的方法均为质量工具。所有的质量工具都以假设为基础。如果偏离了假设，该工具最好的情况下是变成不可预测，最坏的情况则将导致错误的结论。

量具 R&R 研究中最典型的统计假设包括：正态过程、随机和独立的测量、稳定、试验-再试验的准则。当一个或多个假设不成立时（例如，非正态测量过程、操作者偏倚），这个工具和分析最终会变得不稳定、混淆和误导。对产品和过程控制的%GRR 可能会被过高估计。也有同测量系统有关的非统计假设（如校准、可操作性、共同因素和扩张比率、物理定律和常数）。测量的计划者应能识别、控制、纠正或修正更改 MSA 方法，以弥补在测量过程中假设的重大违背。

试验—再试验准则的违背通常是在破坏性或线上测量一项考量，这种测量系统的特性是变化的。测量的计划者要对标准的测量系统研究考虑适当的混合、修改，或其它的技术。

假设有助于对总测量变差做出更大贡献，尤其是当：

- 1) 应用更有效的方法和工具（如 ANOVA、回归分析、试验 设计、概率预测和控制图）。
- 2) 测量精度提高。

应用高精度的测量通常必须为测量系统中的共同因素和热膨胀、变形、蠕变或其它假设，实施计划和偶而的应用，以及纠正。

一质量分析者或测量计划者的最大危害，是假设它们是不重要的或是恒定的，因此经常被忽略。

下表显示了 3 种不同测量情况下 PISMOEA 模型的展示范例。

误差来源	典型生产，自动 MSA	自动化的线上或试验台	校准
P	随机生产件，整个过程范围	生产组件、试验样件、检查标准、加工品	量具、UUT、试验样件、被测物
I	单独的生产量具型式	DCC CMM、试验台	基准量具和设备
S	比率、基准标准或分类；符合“10: 1”比例原则	比率和几何学、参考的试验标准	基准、参考、固有的或一致的、加工品
M	标准操作程序 (S.O.P)，经常开头的，可能被文件化；控制计划	文件化 S.O.P、DCC 程序或自动试验周期	文件化、正式校准程序
O	(2-3) 代表性的，经培训的、经常操作的人	受限的试验操作者、特殊培训与技能	合格的技术员、ISO17025 专业证据
E	稳定的生产和操作条件	经常受控	控制限、优化的，一个主要误差源
A	统计的、通常被忽略	统计的、特定的应用	不能假设的，一种主要误差源
目的	过程控制 (SPC)	产品控制，100% 检验	产品控制，校准公差

特定的误差源的测量变差的影响和贡献的程度根据情况而定。矩阵图、因果分析或失效树诊断将是识别和理解测量变差主要来源的有用工具，进而控制和改进测量系统。

测量系统分析始于理解测量的目的和过程。所有浑沌和不合理的误差来源。测量研究是一个已计划的试验，遵循一些简单的原理：定义重大的误差来源，解决一部分、允许一个或多个控制因素的改变、测量多次、分析结果，以及采取措施。



## 术语

关于其它术语定义，见“统计过程控制”（SPC）参考手册

**5.15 对 6 $\sigma$  相乘系数** 见 VI 页注

**准确度** 一观测值和可接受的参考值之间的一致接近程度。

**方差分析** 通常被用于试验设计（DOE）的一种统计方法（ANOVA），用于分析多个群体中计量型数据，以便比较变差的意义和分析其来源。

**可视解析度** 测量仪器最小增量的大小叫可视解析度。该数值通常广泛的用在公告资料中，以划分测量仪器的等级。数据的分类数可以文字形式（如广告中）来划分测量仪器的分级。数据的分级数可通过把该增量的大小划分类预期的过程分布范围（6 $\sigma$ ）来确定。

注：显示或报告的位数不一定总表示仪器的分辨率。例如，零件的测量值为 29.075、29.080、29.095 等，记录为 5 位数。然而该仪器的分辨率为 0.005 而不是 0.001。

**评价人变差** 在一个稳定环境中应用相同的测量仪器和方法，不同评价人（操作者）对相同零件（被测体）的测量平均值之间的变差。评价人变差（AV）是测量系统变差（误差）的普通原因变差之一，它是由于评价人使用同一测量系统的技能和方法之间的差异所导致的变差。通常将评价人变差设为与测量系统有关的“再现性误差”但这种假设不总是正确的（见再现性术语）。

**偏倚** 测量观测平均值（在重复条件下的测量）与参考值之间的差值；历史上称为准确度。通过一个单值点是否落在测量系统工作范围内来评价和表述偏倚。

**校准** 在规定条件下，建立测量装置和已知基准值和不确定度的可溯源标准之间的关系的一组操作。校准可能也包括通过调整被比较的测量装置的准确度差异而进行的探测、相关性、报告或消除的步骤。

**校准周期** 两次校准间的规定时间总量或一组条件，在此期间，测量装置的校准参数被认定为有效的。

**能力** 以测量系统短期评定为基础的一种测量误差的合成变差（随机的和系统的）的估计。

**置信区间** 期望包括一个参数的真值的值的范围（在希望的概率情况下叫置信水平）。

**控制图** 一种按时间顺序以样本测量为基础的过程特性图形，（这种图形）用于显示过程的行为，识别过程变差的形式，评价稳定性并指示过程方向。

**数据** 一组条件下观察结果的集合，既可以是连续的（一个量值和测量单位）又可以是离散的（属性数据或计数数据如成功/失败、好坏、过/不通过等统计数据）。

<b>设计的试验</b>	一种包含一系列试验统计分析的有计划的研究，在试验中，有目的地改变过程因子并观察结果，以便确定过程变量之间的联系并改进过程。
<b>分辨力</b>	（别名）又称 <b>最小可读单位</b> ，分辨力是测量分辨率、刻度限值或测量装置和标准的最小可探测单位。它是是弄虚作假设计的一个固有特性，并作为测量或分级的单位被报告。数据分级数通常称为“ <b>分辨力比率</b> ”，因为它描述了给定的观察过程变差能可靠地划分为多少级。
<b>明显的</b> 数据分级	能通过测量系统有效分辨率和特定应用于下被观察过程的零件变差可靠地区分开的 <b>数据分级</b> 或分类。见 <b>ndc</b> 。
<b>有效分辨率</b>	考虑整个测量系统变差时数据分级大小叫有效分辨率。基于测量系统变差的置信区间长度来确定该等级的大小。通过把该数据大小划分为预期的过程分布范围能确定数据分级数（ <b>ndc</b> ）。对于有效分辨率，该 <b>ndc</b> 的标准（在 97% 置信水平）估计值为 1.41[PV/GRR]。（见 Wheeler, 1989, 一书中的另一种解释。）
<b>F 比</b>	在选定的置水平上，用于评估随机发生概率的一系列数据的组间均方误差与同组内均方误差之间的数学比率的统计表达。
<b>量具 R&amp;R (GRR)</b>	一个测量系统的重复性和再现性的合成变差的估计。 <b>GRR</b> 变差等于系统内和系统变差之和。
<b>直方图</b>	分组数据的频率的一种图形表示（条形图），用来提供数据分布的直观评价。
<b>受控</b>	只表现出随机、普通原因变差的过程的状态（与无序、指定的或特殊原因变差相反）。只有随机变差的过程操作是统计稳定的。
<b>独立</b>	一个事件或变量的发生对另一个事件或变量发生的概率没有影响。
<b>独立和相同的分布</b>	通常叫“ <b>iid</b> ”。一组同质的数据，这些数据相互独立并随机分布于一个普通分布之中。
<b>交互作用</b>	源于两个或多个重要变量的合成影响或结果，评价人和零件之间具有不可附加性。评价差别依赖于被测零件。
<b>线性</b>	测量系统预期操作范围内偏倚误差值的差别。换句话说，线性表示操作范围内多个和独立的偏倚误差值的相关性。
<b>长期能力</b>	对某个过程长时间内表现的子组内的统计量度。它不同于性能，因为它不包括子组间的变差。
<b>被测体</b>	在规定条件下被测量的特殊数量或对象；对于测量应用一个定义的系列规范。
<b>测量系统</b>	用于量化一个测量单位或确定被测特性性质的仪器或量具、标准、操作、方法、夹具、软件、人员、环境、和条件的集合；用来获得测量的整个过程。

<b>测量系统误差</b>	由于量个偏倚、重复性、再现性、稳定性和线性产生的合成变差。
<b>计量学</b>	测量的科学
<b>ndc</b>	分级数。1.41(PV/GRR)
<b>不可重复性</b>	由于被测体的动态性质决定的对相同样本或部件重复测量的不可能性。
<b>分级数</b>	见 ndc
<b>不受控</b>	表现出混乱的、可指定的或特殊原因变差的过程的状态。不受控的过程即统计不稳定。
<b>零件间变差</b>	与测量系统分析有关，对于一个稳定过程零件变差（PV）代表预期的不同零件和不同时间的变差。
<b>性能</b>	以测量系统长期评价为基础的测量误差（随机的和系统的）合成变差的估计，包括所有随时间变化的显著的和可确定的变差源。
<b>精密度</b>	测量系统在操作范围内（容量、范围和时间）的分辨力、敏感性和重复性的净效果。在一些组织中，精密度和重复性具有互换性。事实上，精密度最经常用于描述测量范围内的预期重复测量变差，这个范围可以是容量和时间。通常建议使用比术语“精密”更具有描述性的术语。
<b>概率</b>	以已收集数据的特定分布为基础的，描述特定事件发生机会的一种估计（用比例或分数）。概率估计值范围从0（不可能事件）到1（必然事件）。一组条件或原因共同作用产生某种结果。
<b>过程控制</b>	一种运行状态，将测量目的和决定准则应用迁实时生产以评估过程稳定性和测量体或评估自然过程变差的性质。测量结果显示过程或者是稳定和“受控”，或者是“不受控”。
<b>产品控制</b>	一种运行状态，将测量目的和决定准则应用于评价测量体或评价特性符合某规范。测量结果显示过程或是“在公差内”或者是“在公差外”。

<b>基准值</b>	<p>轴承认的一个被测体的数值，作为一致同意的用于进行比较的基准或标准样本：</p> <ul style="list-style-type: none"> <li>● 一个基于科学原理的理论值或确定值；</li> <li>● 一个基于某国家或国际组织的指定值；</li> <li>● 一个基于某科学或工程组织主持的合作试验工作产生的一致同意值；</li> <li>● 对于具体用途，采用接受的参考方法获得的一个同意值。</li> </ul> <p>该值包括特定数量的定义，并为其它已知目的的自然接受，有时是按惯例被接受。</p> <p>注：与基准值同义使用的其它术语：</p> <ul style="list-style-type: none"> <li>已接受的基准值</li> <li>已接受值</li> <li>惯用值</li> <li>惯用真值</li> <li>指定值</li> <li>最佳估计值</li> <li>标准值</li> <li>标准测量</li> </ul>
<b>回归分析</b>	<p>两个或多个变量之间的关系的统计研究。确定两个或多个变量间数学关系的一种计算。</p>
<b>重复性</b>	<p>在确定的测量条件下，来源于连续试验的普通原因随机变差。通常指设备变差(EV)尽管这是一个误导。当测量条件固定和已定义时，即确定零件、仪器标准、方法、操作者、环境和假设条件，适合重复也包括在特定测量误差模型下条件下的所有内部变差。</p>
<b>可重复性</b>	<p>对相同样件或部件进行重复测量的能力，被测体或测量环境没有明显的物理变化。</p>
<b>重复</b>	<p>重复性（相同的）条件下的多次实验。</p>
<b>再现性</b>	<p>测量过程中由于正常条件改变所产生的测量均值的变差。一般来说，它被定义为在一个稳定环境下，应用相同的测量仪器和方法，相同零件（被测体）不同评价人（操作者）之间测量值均值的变差。这种情况对受操作者技能影响的手动仪器常常是正确的，然而，对于操作者不是主要变差源的测量过程（如自动系统）则不正确的。由于这个原因，再现性指的是测量系统之间和测量条件之间的均值变差。</p>

<b>分辨率</b>	<p>可用作<b>测量分辨率</b>或<b>有效分辨率</b>。测量系统探测并如实显示被测特性微小变化的能力。（参见分辨力）</p> <p>如果对与标准零件之差小于 <math>\delta</math> 的任何零件的指示值与标准零件指示值概率相等，则测量系统分辨率为 <math>\delta</math>。测量系统的分辨率受测量仪器以及整个测量系统其它变差源的影响。</p>
<b>散点图</b>	数据的 X-Y 坐标图，用于评估两个变量之间的关系。
<b>敏感性</b>	<p>导致一个测量装置产生可探测（可辨别）输出信号的最小输入信号。一个仪器应至少和其分辨力单位同样敏感。敏感性是通过固有量具的设计与质量、服务期内维护和操作条件确定。敏感性是用测量单位报告的。</p>
<b>显著水平</b>	被选择用来测试随机输出概率的一个统计水平，也同风险有关，表示为 $\alpha$ 风险，代表一个决定出错的概率。
<b>稳定性</b>	既指测量过程的统计稳定性又指随时间变化的测量稳定性。两者对测量系统预期用途都是重要的。统计稳定性包含一个可预测的、潜在的测量过程，该过程在普通原因变差（受控）条件下运行。测量稳定性（别名漂移）代表测量系统在运行周期（时间）内对测量标准或基准的必要的符合程度。
<b>容差（公差）</b>	为了维持配合、形式和功能，与标准值或公称值相比允许的偏差。
<b>不确定度</b>	同测量结果有关的一个参数，代表数值的分散特性，此数值归结于被测体（VIM）是合理的。在给定的置信水平内，对一个测量结果的指定范围描述，限值期望包含真实测量结果。不确定度是一个测量可靠性的量化表述。
<b>单峰</b>	具有一种模式的一组邻近的数据。

# 样表

允许读者复制本节的表格。

本节中的表格表述了用于 GRR 的数据收集和报告格式。  
还有其他类型的可能包含相同信息和获得同样结果的格式。  
它们不是唯一的格式。

量具重复性和再现性数据收集表

评价人/ 试验#	零件										均值
	1	2	3	4	5	6	7	8	9	10	
A 1											
2											
3											
均值											$\bar{X}_a =$
极差											$\bar{R}_a =$
B 1											
2											
3											
均值											$\bar{X}_b =$
极差											$\bar{R}_b =$
C 1											
2											
3											
均值											$\bar{X}_c =$
极差											$\bar{R}_c =$
零件均值											$\bar{\bar{X}} =$ $R_p =$
$([\bar{R}_a = \quad] + [\bar{R}_b = \quad] + [\bar{R}_c = \quad]) / [\text{评价人数} \quad] =$											$\bar{R} =$
$\bar{X}_{DIFE} = [\text{Max} \bar{X} = \quad] - [\text{Min} \bar{X} = \quad] =$											$\bar{X}_{DIFE}$
$*UCL_R = [\bar{R} = \quad] \times [D_4 = \quad] =$											
<p>2 次试验 <math>D_4=3.27</math>, 3 次试验 <math>D_4=2.58</math>。UCL<sub>R</sub> 代表了单个极差的控制限。圈出那些超过控制限的点，识别原因并纠正。使用与开始时相同的评价人及单位重复这些读数，或者放弃某些值，从保留的观察值重新平均，重新计算 <math>\bar{R}</math> 和控制限。</p> <p>注： _____</p>											

### 量具重复性和再现性报告

零件号和名称:	量具名称:	日期:										
特性:	量具号:	完成人:										
规范:	量具类型:											
$\bar{R} =$	$X_{DIFF} =$	$R_p =$										
测量单元分析		% 总变差 (TV)										
重复性—设备变差 (EV) $EV = \bar{R} \times K_1$ $= \text{---} \times \text{---}$ $= \text{---}$		$\%EV = 100[EV/TV]$ $= 100[\text{---} / \text{---}]$ $= \text{---} \%$										
	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%; text-align: center;"><math>K_1</math></td> </tr> <tr> <td style="text-align: center;">3</td> <td style="text-align: center;">0.8862</td> </tr> <tr> <td style="text-align: center;">3</td> <td style="text-align: center;">0.5908</td> </tr> </table>			$K_1$	3	0.8862	3	0.5908				
	$K_1$											
3	0.8862											
3	0.5908											
再现性—设备变差 (AV) $AV = \sqrt{(\bar{X}_{DIFF} \times K_2)^2 - (EV^2 / (nr))}$ $= \sqrt{(\text{---})^2 - (\text{---}^2 / (\text{---} \times \text{---}))}$ $= \text{---}$		$\%AV = 100[AV/TV]$ $= 100[\text{---} / \text{---}]$ $= \text{---} \%$										
	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%;"></td> <td style="width: 33%; text-align: center;">2</td> <td style="width: 33%; text-align: center;">3</td> </tr> <tr> <td style="text-align: center;"><math>K_2</math></td> <td style="text-align: center;">0.7071</td> <td style="text-align: center;">0.5231</td> </tr> </table>			2	3	$K_2$	0.7071	0.5231				
	2		3									
$K_2$	0.7071	0.5231										
n =	r =											
重复性和再现性(GRR) $GRR = \sqrt{EV^2 + AV^2}$ $= \sqrt{(\text{---})^2 + (\text{---})^2}$ $= \text{---}$		$\%GRR = 100[GRR/TV]$ $= 100[\text{---} / \text{---}]$ $= \text{---} \%$										
	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%; text-align: center;"><math>K_3</math></td> </tr> <tr> <td style="text-align: center;">2</td> <td style="text-align: center;">0.7071</td> </tr> </table>			$K_3$	2	0.7071						
	$K_3$											
2	0.7071											
零件变差(PV) $PV = R_p \times K_3$ $= \text{---} \times \text{---}$ $= \underline{1.10456}$	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; text-align: center;">3</td> <td style="width: 50%; text-align: center;">0.5231</td> </tr> <tr> <td style="text-align: center;">4</td> <td style="text-align: center;">0.4467</td> </tr> <tr> <td style="text-align: center;">5</td> <td style="text-align: center;">0.4030</td> </tr> </table>	3	0.5231	4	0.4467	5	0.4030	$\%PV = 100[PV/TV]$ $= 100[\frac{1.1045}{1.14610}]$ $= \underline{96.38} \%$				
3	0.5231											
4	0.4467											
5	0.4030											
总变差(TV) $TV = \sqrt{GRR^2 + PV^2}$ $= \sqrt{(\text{---})^2 + (\text{---})^2}$ $= \text{---}$	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; text-align: center;">6</td> <td style="width: 50%; text-align: center;">0.3742</td> </tr> <tr> <td style="text-align: center;">7</td> <td style="text-align: center;">0.3534</td> </tr> <tr> <td style="text-align: center;">8</td> <td style="text-align: center;">0.3375</td> </tr> <tr> <td style="text-align: center;">9</td> <td style="text-align: center;">0.3249</td> </tr> <tr> <td style="text-align: center;">10</td> <td style="text-align: center;">0.3146</td> </tr> </table>	6	0.3742	7	0.3534	8	0.3375	9	0.3249	10	0.3146	$ndc = 1.41(\frac{PV}{GRR})$ $= 1.41(\text{---} / \text{---})$ $= \text{---}$
6	0.3742											
7	0.3534											
8	0.3375											
9	0.3249											
10	0.3146											
表格中理论和常数信息参见 MSA 参考手册,第三版												



