

中华人民共和国国家标准

GB/T 6379.2—2004/ISO 5725-2:1994

部分代替 GB/T 6379—1986

GB/T 11792—1989

测量方法与结果的准确度

(正确度与精密度)

第2部分:确定标准测量方法重复性 与再现性的基本方法

Accuracy (trueness and precision) of measurement methods and results—

Part 2: Basic method for the determination of repeatability and
reproducibility of a standard measurement method

(ISO 5725-2:1994, IDT)

2004-06-02 发布

2005-01-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会

发布

前 言

GB/T 6379《测量方法与结果的准确度(正确度与精确度)》分为六部分,其预期结构及对应的国际标准为:

- 第1部分:总则与定义(ISO 5725-1:1994, IDT)
- 第2部分:确定标准测量方法的重复性和再现性的基本方法(ISO 5725-2:1994, IDT)
- 第3部分:标准测量方法精密度的中间度量(对应 ISO 5725-3:1994)
- 第4部分:确定标准测量方法正确度的基本方法(对应 ISO 5725-4:1994)
- 第5部分:确定标准测量方法正确度的可替代方法(对应 ISO 5725-5:1998)
- 第6部分:准确度值的实际应用(对应 ISO 5725-6:1994)

本部分为 GB/T 6379 的第2部分。

GB/T 6379 的本部分等同采用国际标准 ISO 5725-2:1994《测量方法与结果的准确度(正确度与精确度) 第2部分:确定标准测量方法重复性和再现性的基本方法》及 ISO 于 2002-05-15 发布的对 1994 版 ISO 5725-2 的技术修改单。

GB/T 6379 第1部分至第6部分作为一个整体代替 GB/T 6379—1986 和 GB/T 11792—1989。标准中将原精密度加以扩展,增加了正确度;除重复性条件和再现性条件外,增加了中间精密度条件。

本部分的附录 A 和附录 B 为规范性附录,附录 C 为资料性附录。

本部分由中国标准化研究院提出。

本部分由全国统计方法与应用标准化技术委员会归口。

本部分起草单位:中国标准化研究院、中国科学院数学与系统科学研究院、辽宁出入境检验检疫局、广州出入境检验检疫局。

本部分主要起草人:于振凡、冯士雍、刘文、姜健、丁文兴、王斗文、肖惠、李成明。

本部分于 2004 年首次发布。

引 言

0.1 GB/T 6379 用两个术语“正确度”与“精密度”来描述一种测量方法的准确度。正确度指大量测试结果的(算术)平均数与真值或接受参照值之间的一致程度;而精密度指测试结果之间的一致程度。

0.2 GB/T 6379.1 中对上述诸量给出了一般性的考虑,在 GB/T 6379 本部分中不再重复。GB/T 6379.1 应与 GB/T 6379 所有其他部分(包括本部分)结合起来读,因为 GB/T 6379.1 给出了基本定义和总则。

0.3 GB/T 6379 本部分只考虑重复性标准差和再现性标准差的估计。虽然其他类型的试验(如分割水平试验)也可用于某些情形的精密度估计,GB/T 6379 本部分中并没有涉及这方面内容,这将在 ISO 5725-5 中进行讨论。GB/T 6379 本部分也没有考虑精密度两个主要度量之间的任何其他中间度量问题,这将是 GB/T 6379.3 的内容。

0.4 在某些情况下,为进行精密度估计得到的试验数据也可用于估计正确度。正确度的估计也不在本部分中给予考虑,有关正确度估计的所有内容将是 GB/T 6379.4 的目标。

测量方法与结果的准确度 (正确度与精密度)

第2部分:确定标准测量方法重复性 与再现性的基本方法

1 范围

1.1 GB/T 6379 的本部分

——给出了一些通过协同试验室间试验获得测量方法精密度的数值估计的试验设计中应遵循的一般原则。

——提供估计测量方法精密度所常用的基本方法的详细而实用的说明。

——为所有关心精密度估计的试验设计、执行和结果分析人员提供指南。

注1:为特定目的而对基本方法进行某些修正改变的内容在 GB/T 6379 其他部分中给出。

附录 B 提供了一些通过试验估计测量方法精密度的实际例子。

1.2 GB/T 6379 本部分所涉及的测量方法特指对连续量进行测量,并且每次只取一个测量值作为测试结果的测量方法,尽管这个值可能是一组观测值的计算结果。

1.3 假定在精密度试验的设计和执行过程中,已经遵循了在 GB/T 6379.1 中给定的所有原则。基本方法是在每一个实验室使用相同数目的测试结果,并且每一个实验室都对测试样本的相同水平进行分析,即进行平衡均匀水平的试验。基本方法适用于那些标准化的且在很多实验室常规使用的程序。

注2:标准附录部分给出了几个实际例子来说明平衡一致的测试结果。在一个例子中,单元中的重复测试数有变化(非均衡设计);在另一个例子中一些数据缺失了。这是因为一个均衡设计的试验最终可能变成非均衡的。也考虑了歧离值和离群值。

1.4 GB/T 6379.1—2004 第5章中的统计模型被用作解释和分析测试结果的基本模型,且认为测试结果是近似服从正态分布的。

1.5 GB/T 6379 本部分估计测量方法精密度的基本方法用于下列情形:

- a) 需要确定在 GB/T 6379.1 中定义的重复性和再现性标准差时;
- b) 当使用的物料是匀质的,或不均匀性的影响能包括在精密度值中时;
- c) 当可以采用平衡均匀水平设计时。

1.6 对尚未标准化的或非常规使用的测量方法进行初步精密度估计时也可以使用同样的方法。

2 规范性引用文件

下列文件中的条款通过 GB/T 6379 本部分的引用而成为本部分的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版本均不适用于本部分,然而,鼓励根据本部分达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本部分。

ISO 3534-1:1993 统计学 词汇和符号 第1部分:概率和一般统计术语

GB/T 6379.1—2004 测量方法与结果的准确度(正确度与精密度) 第1部分:总则与定义

3 定义

ISO 3534-1 和 GB/T 6379.1 中给出的定义在 GB/T 6379 本部分中仍适用。

GB/T 6379 使用的符号由附录 A 给出。

4 基本模型中的参数估计

4.1 GB/T 6379 本部分给出的程序是建立在 GB/T 6379.1—2004 第 5 章的统计模型基础上的,这些程序在 GB/T 6379.1—2004 的 1.2 中进行了详细的阐述。特别是,这些程序是基于 GB/T 6379.1—2004 第 5 章中的式(2)到式(6)。

基本模型是:

$$y = m + B + e$$

其中,对给定的受试物料:

m ——总平均值(期望);

B ——重复性条件下的偏倚的实验室分量;

e ——在重复性条件下每一次测量产生的随机误差。

4.2 GB/T 6379.1—2004 第 5 章中的式(2)到式(6)表示的是所考虑总体标准差的真值。实际情况中,这些标准差的确切值是未知的,精密度的估计值通过从全体实验室组成的总体抽取少量的实验室来获得的。而在这些实验室内,该估计值由所有可能测试结果的一个小样本获得。

4.3 在统计实践中,如果标准差的真值 σ 未知,则以样本进行估计并替代,此时,符号 σ 用 s 代替, s 表示 σ 的估计值。下列估计值可根据 GB/T 6379.1—2004 的式(2)到式(6)得出:

s_L^2 :实验室间方差的估计值;

s_W^2 :实验室内方差的估计值;

s_r^2, s_W^2 的算术平均值,并且是重复性方差的估计值;这个算术平均值是在剔除了离群值后对所有参与准确度试验的实验室计算的。

s_R^2 :再现性方差的估计值:

$$s_R^2 = s_L^2 + s_r^2 \dots\dots\dots (1)$$

5 对精密度试验的要求

5.1 试验安排

5.1.1 在用基本方法进行试验安排时,取自 q 批物料的样本分别代表 q 个不同测试水平,被分到 p 个实验室,每一个实验室都在重复性条件下对每一水平得到同样 n 次重复测试结果。这种试验称为平衡均匀水平试验。

5.1.2 这些测量工作应在如下规则下组织进行:

- a) 任何设备的预检应按标准方法中的规定进行;
- b) 同一水平中一组 n 次测量应该在重复性条件下进行,即在短暂的时间间隔内,由同一操作员测量;除非是作为整个测量过程的一个环节,测量过程中间不允许对设备进行任何的重新校准。
- c) 一组 n 次测试要求在重复性条件下独立地进行是十分重要的,就像是在对 n 种不同的物料进行的 n 个测试。然而,事实上,操作员会知道他是对同一物料进行测试。应在说明书中强调的是,测试的整个意图就是要考察在实际测试中测试结果能发生多大的变化。尽管有这样的提示,为避免前面的测试结果对随后的测试产生影响,从而影响重复性方差,可考虑在全部 q 个水平,每个水平上要求 n 个独立测试的样本,混合进行编号,使得操作员不知道所进行的测试是哪个水平的。不过,这样的程序也可能产生另一个问题,即能否保证重复性条件适用于这些重复的测试。只有当所有 qn 个测量可以在一个很短的时间内完成时,上述条件才能得到保证。
- d) 没有必要要求所有 q 组的 n 次测量都严格在一个很短的时间内进行;不同组的测量可以不在同日内进行;
- e) 所有 q 个水平的测量都将由同一个操作员做出,此外,在给定水平上做出的 n 个测量要自始至

终使用同一设备；

- f) 如果在测量过程中一个操作员因故不能完成全部测量,那么可以由另一个操作员继续剩下的测量,只要这个人员变更不是发生在同一水平同一组的 n 个测量上,而是发生在 q 组中的两个不同组上。任何这样的人员变更都要随测试结果一起上报。
- g) 应该给出一个时间限制,所有的测量应该在该时间区间内完成。把该时间限制在收到样本的日期和测量完成的日期之间；
- h) 所有的受试样本都应该用标签标明测试名称并对样本进行编号。

5.1.3 对 5.1.2 和 GB/T 6379 本部分的其余章节中“操作员”加以说明。对于某些测量,事实上可能由一组操作员进行,每一操作员执行测量程序的某一规定部分。在此情况下,这一组操作员将统一看作“操作员”,这一组中出现任何人员的变更都将被看成是不同的“操作员”。

5.1.4 在商业实践中,对测试结果的修约可能做得很粗。但在精密度试验中,测试结果要比标准方法中规定的有效数字位数至少多一位。如果该方法没有规定有效数字位数,那么修约的误差不能超过重复性标准差估计值的 $1/2$ 。当精密度依赖于水平 m 时,对于不同的水平就要有不同的修约程度。

5.2 实验室征集

5.2.1 在 GB/T 6379.1—2004 的 6.3 中给出了关于参与实验室间协同试验的实验室征集工作的一般原则。在征集所需数目的协同实验室时,要明确规定这些实验室的条件。图 1 中给出了一个实验室调查征集的例子。

5.2.2 一个“实验室”在 GB/T 6379 本部分中被认为是操作员、设备和测试场所的一个组合,一个测试场所或通常意义的一个实验室可以产生几个“实验室”,只要它能够为几个操作员提供独立的仪器设备和测试场地。

5.3 物料准备

5.3.1 在 GB/T 6379.1—2004 的 6.1 中给出了精密度试验中选择物料时需要考虑的要点。

5.3.2 在决定试验所需的物料数量时,应该考虑到在获得某些测试结果时会出现偶然的洒出和称量误差,从而需用到额外的物料。需要准备的物料数量应当足以满足测试之用,并且允许适当的储备。

实验室间协同研究调查表	
测量方法名称(附相关资料的复印件)	
1. 本实验室将参加这个标准测量方法精密度试验	
是 ☐	不是 ☐ (在相应的方框内打勾)
2. 作为参与者,我们理解到:	
a) 所有必要的仪器设备、化学品和在该方法中规定的其他条件在项目开始时都必须在我们实验室备齐;	
b) 规定中的时间要求,包括开始日期,测试样品的顺序和项目的完成时间都必须严格遵照执行;	
c) 必须严格按方法要求进行试验;	
d) 样本必须按照说明书规定处置;	
e) 测量必须由一位合格的操作员完成。	
我们已经认真地研究了该测量方法并且也已对我们的能力和设备进行了合理的评估,我们认为有足够的实力参加这个方法的协同试验。	
3. 意见	
(签名)	
(公司或实验室)	

图 1 实验室间协同研究调查表

5.3.3 应考虑在得到正式的测试结果之前一些实验室为了熟悉测量方法而获得某些初步测试结果是否可取,如果可取,那么也应考虑是否应该提供额外的物料(非精密度试验样本)。

5.3.4 当一种物料必须要进行匀质化时,应对该种物料以最合适的方式进行匀质化。当要进行测试的物料不是匀质时,就要以该方法中规定的方式准备样本,这是很重要的,最好对每个水平都用不同批的商业物料。对于不稳定的物料,应给出特殊的储藏和处置说明书。

5.3.5 如果容器一旦被打开物料就有变质的危险(例如被氧化,损失挥发或吸湿物料),那么对于每一水平下的样本,应对每个试验室使用 n 个不同的容器。在物料不稳定的情况下,应给出特殊的储藏和处置说明书。应该采取一些预防措施来确保样本直到进行测量时相同。如果要测量的物料是由不同相对密度的粉状物料混合而成或由不同大小的颗粒组成的,那么由于震动可能会产生分离(例如在运输过程中),因此需要特别注意。当受试样本可能与空气发生反应时,样品可以被封在被抽空或者用惰性气体填充的玻璃瓶内。对于食品或血样这样的易变质的物料,有必要将其以冷冻状态送到参与的实验室,并对其融化程序进行详细的说明。

6 参与精密度试验的人员

注3:不同的实验室其操作方法不尽相同。因此,本章的内容仅作为一个指南,在特定情况下可作适当修改。

6.1 领导小组

6.1.1 领导小组宜由熟悉该测量方法及其应用的专家组成。

6.1.2 领导小组的任务是:

- a) 计划和协调整试验;
- b) 决定需要的实验室数量、水平和要求的测量数,以及要求的有效数字位数;
- c) 指定其中某位成员承担统计方面的职责(见 6.2);
- d) 指定其中一位成员为执行负责人(见 6.3);
- e) 考虑给每个实验室的测量负责人下发除了标准测量方法以外的操作说明书;
- f) 决定是否允许某些操作员进行少量的非正式测量,以便在间歇很长时间后(这些测量结果不应作为协同试验的正式样本)重获测量方法方面的经验;
- g) 测试结果分析完成后,讨论统计分析报告;
- h) 确定重复性标准差和再现性标准差的最终值;
- i) 决定是否需要就改进测量方法标准及对那些测试结果被作为离群值而拒绝的实验室采取进一步的措施。

6.2 统计专家的职责

领导小组中至少有一个成员应具有统计设计和试验分析方面的经验。他的任务是:

- a) 用专业知识进行试验设计;
- b) 对数据进行分析;
- c) 按 7.7.1 中的规定向领导小组提交一份报告。

6.3 执行负责人的职责

6.3.1 把试验实际的组织工作委托给某个实验室。领导小组任命该实验室的一名成员为执行负责人,对此工作负全责。

6.3.2 执行负责人的任务是:

- a) 征集必要数目的协同实验室,并且负责任命每个实验室的测量负责人;
- b) 组织和监管测试物料、样本的准备以及样本的分配;对每个水平,应该预留足够量的备用物料;
- c) 起草涵盖 5.1.2 中 a) 至 h) 各项要点的操作说明书,将说明书尽早地提前分发给各实验室测量负责人,以便他们能对其提出意见,确保所选的操作员在常规操作中能正确地进行测量;

- d) 设计适当的表格,以便操作员用于工作记录、测量负责人用于报告测试结果的有效数字位数;(表格可以包括操作员的姓名、收到和测量样本的日期、所使用的设备和其他有关的信息等);
- e) 处理各实验室在测量操作中出现的問題;
- f) 关注试验的进度,使试验按规定日程进行;
- g) 收集数据表并把它们提交给统计专家。

6.4 测量负责人

6.4.1 每个参与试验的实验室应指定一名成员负责实际测量的组织、按执行负责人的指令工作并报告测试结果。

6.4.2 测量负责人的任务是:

- a) 确保所选的操作员在日常操作中能正确地进行测量;
- b) 按执行负责人的指令把样本分发给操作员(必要时还要为熟悉试验操作提供物料);
- c) 对测量的执行进行监管(测量负责人不应参与测量操作);
- d) 确保操作员进行规定次数的测量;
- e) 确保测量工作按时间进度进行;
- f) 收集测试结果,要求结果记录的小数位数与要求一致,以及测试中遇到的任何困难、异常现象和操作员反映的意见。

6.4.3 每个实验室的测量负责人应撰写一份包括下面信息的全面报告:

- a) 原始测试结果,由操作员以清晰字迹记录在所提供的表格上,而不要转录或打印(计算机或测试机器打印输出的结果除外);
- b) 最初的观测值或读数(当测试结果由这些读数计算得出时),由操作员以清晰字迹记录在所提供的表格上,而不要转录或打印;
- c) 操作员提出的关于测量方法标准方面的意见;
- d) 在测量期间发生的任何非常规或干扰的信息,包括可能发生的操作员变更,指明哪位操作员做了哪些测量,以及对任何数据缺失原因的说明;
- e) 样本收到的日期;
- f) 每个样本被测量的日期;
- g) 使用的相关设备信息;
- h) 其他有关的信息。

6.5 操作员

6.5.1 在每个实验室中,测量应该由一个选定的操作员完成,该操作员是在通常操作中可能执行该测量任务的操作员代表。

6.5.2 因为试验的目的是对全体使用该标准测量方法的操作员确定标准测量方法的精密度,因此一般不宜给操作员以拓展测量方法标准的权利。然而,也应该对操作员指出,测试的目的之一是发现测试结果在实际中的变化,这样他们就不会对不一致的测量结果进行丢弃或重测。

6.5.3 尽管操作员通常没有对标准测量方法进行补充性修订的任务,但是也应鼓励他们为标准做出评价,尤其是指出标准中的说明是否足够明确而不模糊。

6.5.4 操作员的任务是:

- a) 根据标准测量方法实施测量;
- b) 报告测试中遇到的异常现象和困难;报告一个错误要比调整测试结果更为重要,因为缺失一两个测试结果不会毁坏整个试验,多数情况下反而反映了测量标准本身的不足。
- c) 为评价标准中的说明是否合适;操作员应在遇到的任何不能按试验说明进行测试的情况时随时报告,因为这也反映了标准本身的不足。

7 精密度试验的统计分析

7.1 初步考虑

7.1.1 数据的分析是一个统计问题,应由统计专家来解决,它包括以下三个相继的步骤:

- a) 对数据进行检查,以判别和处理离群值或其他不规则数据,并检验模型的合适性;
- b) 对每个水平分别计算精密度和平均值的初始值;
- c) 确定精密度和平均值的最初值,且在分析表明精密度和水平 m 之间可能存在某种关系时,建立它们之间的关系。

7.1.2 对每个水平,首先计算以下诸量的估计值:

- 重复性方差: s_r^2
- 实验室间方差: s_L^2
- 再现性方差: $s_R^2 = s_r^2 + s_L^2$
- 平均值 m

7.1.3 统计分析包括对离群值的统计检验的系统应用,在文献中有许多方法可以用于 GB/T 6379 本部分。从实际应用考虑,在 7.3 中将这方法选择整理成几种方法。

7.2 结果列表和所用记号

7.2.1 单元

一个实验室和一个水平的组合称为精密度试验的一个单元。理想的情况是,一项有 p 个实验室和 q 个水平的试验,列成 pq 个单元的表,每个单元包含 n 次重复测试结果,以此来计算重复性标准差和再现性标准差。然而,由于多余数据、缺失数据和离群值的发生,这种理想情况在实际中并不总是能够得到的。

7.2.2 多余数据

有时一个实验室可能进行且报告了多于正式规定的 n 个测试结果。在此情形下,测量负责人应报告为什么会这样? 哪些是正确的测试结果? 如果答案是这些测试结果都是同样有效的,则宜在这些测试结果中随机抽取原定数量的数据用于分析。

7.2.3 缺失数据

另一种情形是,一些测试结果可能缺失,例如,因为样本的丢失或在测量时操作的失误。在 7.1 中推荐的分析程序是对完全空白的单元简单地将其忽略,而对部分空白的单元则通过标准计算程序给予考虑。

7.2.4 离群值

离群值是原始测试结果或由此生成的一些数值,与其他测试结果或同样产生的其他数值相差很大,不一致。经验告诉我们,离群值不能完全避免,须与缺失数据一样作类似处理。

7.2.5 离群实验室

当某个实验室在几个不同水平出现无法解释的非正常测试结果,在所测试水平下,实验室内方差和(或)系统误差过大时,可将它作为离群实验室。有理由舍弃离群实验室的部分或全部数据。

GB/T 6379 本部分没有提供如何判断可疑的离群实验室的统计检验程序。对此应由统计专家做出初步决定,但是所有被除外的实验室都应该报告给领导小组以便采取进一步的行动。

7.2.6 错误数据

有明显错误的的数据应进行核查并予以更正或删除。

7.2.7 平衡均匀水平测试结果

理想的情况是对 p 个实验室(编号为 $i=1,2,\dots,p$), q 个水平(编号为 $j=1,2,\dots,q$),每个水平都重复 n 次测试的情形,总共获得 pqn 个测试结果。由于数据缺失(7.2.3)、离群的测试结果(7.2.4)、离群实验室(7.2.5)或错误数据(7.2.6)的存在,这种理想的情况并不总能得到。在这些情况下,在 7.2.8 到

7.2.10 中的记号和 7.4 中的程序允许测试结果数不全不同。图 2 给出用于统计分析的推荐的原始数据的列表格式,为方便起见,分别简称为 A 表、B 表和 C 表。

A 表 原始数据整理的推荐格式									
实验室	水 平								
	1	2	..	..	j	..	..	q-1	q
1									
2									
..									
..					..				
..					..				
i					y_{ijk}				
..					..				
..									
p									

B 表 单元平均值整理的推荐格式									
实验室	水 平								
	1	2	..	..	j	..	..	q-1	q
1									
2									
..									
i					$\bar{y}_{ij}$				
..									
p									

C 表 单元内离散度的推荐格式									
实验室	水 平								
	1	2	..	..	j	..	..	q-1	q
1									
2									
..									
i					s_{ij}				
..									
p									

图 2 分析结果整理的推荐格式

7.2.8 原始测试结果

图 2 的 A 表中，
 n_{ij} 是第 i 个实验室在水平 j 这个单元的测试结果数；
 y_{ijk} 是该单元第 k 个测试结果 ($k=1,2\cdots n_{ij}$)
 p_j 是 j 水平至少有一个测试结果的实验室数(在剔除了所有离群值和错误的测试结果后)。

7.2.9 单元平均值(图 2 的 B 表)

由 A 表按下式计算单元平均值：

$$\bar{y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \quad \dots\dots\dots (2)$$

单元平均值应比 A 表中的测试结果有效数字多一位。

7.2.10 单元离散度(图 2 的 C 表)

由 A 表(参见 7.2.8)和 B 表(参见 7.2.9)按下面的公式计算单元离散度:

一般情况,使用单元内标准差,即

$$s_{ij} = \sqrt{\frac{1}{n_{ij}-1} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2} \quad \dots\dots\dots (3)$$

上式等价于

$$s_{ij} = \sqrt{\frac{1}{n_{ij}-1} \left[\sum_{k=1}^{n_{ij}} (y_{ijk})^2 - \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{ij}} y_{ijk} \right)^2 \right]} \quad \dots\dots\dots (4)$$

在使用上述公式计算时,应注意在计算过程中保留足够的有效数字位数,每个中间值需要保留的位数应是原始数据的两倍。

注 4: 如果单元 ij 只包含两个测试结果,单元内标准差即为:

$$s_{ij} = |y_{ij1} - y_{ij2}| / \sqrt{2} \quad \dots\dots\dots (5)$$

因此,为简单起见,若所有单元都只包含两个测试结果,则可用绝对差代替标准差。

标准差应该比 A 表中的结果的有效数字多一位。

若 n_{ij} 小于 2,在 C 表中插入符号“—”。

7.2.11 经更正或被剔除的数据

因为一些数据根据 7.1.3, 7.3.3 和 7.3.4 中提到的检验可能经过更正或予以剔除,因此用于最后确定精密度和平均值的 y_{ijk} , n_{ij} 及 p_j 可能与在图 2 中记录的 A 表, B 表和 C 表中的原始测试结果不同。所以在报告精密度和正确度的最终值时,如果有经过更正或剔除的数据应予以指出。

7.3 测试结果的一致性和离群值检查

参见参考文献 3。

根据对多个水平获得的数据,即可对重复性标准差和再现性标准差进行估计。由于个别实验室或数据可能与其他实验室或其他数据明显不一致,从而影响估计,必须对这些数值进行检查。为此介绍以下两种方法:

- a) 检验一致性的图方法;
- b) 检验离群值的数值方法。

7.3.1 检验一致性的图方法

该方法需用到称为曼德尔的 h 统计量和 k 统计量的两种度量。除用来描述测量方法的变异外,这两个统计量对实验室评定也是有用的。

7.3.1.1 对每个实验室的每个水平,计算实验室间的一致性统计量 h ,方法是用单元对平均值的离差(单元平均值减去该水平的总平均值)除以单元平均值的标准差:

$$h_{ij} = \frac{\bar{y}_{ij} - \bar{\bar{y}}_j}{\sqrt{\frac{1}{p_j-1} \sum_{i=1}^{p_j} (\bar{y}_{ij} - \bar{\bar{y}}_j)^2}} \quad \dots\dots\dots (6)$$

上式中, $\bar{y}_{ij}$ 见 7.2.9, $\bar{\bar{y}}_j$ 见 7.4.4。

将 h_{ij} 的数值,按实验室顺序,以每个实验室的不同水平为一组描点作图(参见 B.7,称为 h 图)。

7.3.1.2 对每个实验室 i ,计算实验室内的一致性统计量 k ,方法是先对每个水平 j 计算联合单元内标准差:

$$\sqrt{\frac{\sum s_{ij}^2}{p_j}}$$

然后对每个实验室的每个水平内计算：

$$k_{ij} = \frac{s_{ij} \sqrt{p_j}}{\sqrt{\sum s_{ij}^2}} \dots\dots\dots (7)$$

将 k_{ij} 的数值,按实验室顺序,以每个实验室的不同水平为一组,描点作图(参见图 B.8,称为 k 图)。

7.3.1.3 检查 h 与 k 图可以发现是否有这样实验室,它的测试结果与所考察的其他实验室明显不同。这里的不同表示为单元内变异一致的高或低,或者单元平均值在许多水平上皆为最高或最低。若发生此种情况,应与该实验室接触,探究造成此类不同行为的原因,根据调查结果统计专家可:

- a) 暂时保留该实验室的数据;
- b) 要求实验室重新进行测量(如果可行);
- c) 剔除该实验室的数据。

7.3.1.4 h 图有不同的模式。对试验的不同水平,实验室的 h 值可正可负。一个实验室的 h 值可能皆为正值,或皆为负值,取负值的实验室数与取正值的实验室数大致相等。虽然上述第二种模式表明有共同的实验室偏倚来源的可能,但这两种模式都是正常的,不需要做特别的检查。另一方面,若有一个实验室的 h 值皆取同一符号(正或负),而所有其他实验室的 h 值皆取另一种符号,就需要查找原因。类似的,若一个实验室的 h 值比较极端,且与试验水平有系统的依赖关系,则也需查找原因。在 h 图上按 8.3 中(表 6 与表 7)的临界值画出的临界线,可用于考察数据的行为模式。

7.3.1.5 如果一个实验室的 k 图上的多个点值都很大,就要查找原因,这表明该实验室的重复性比其他实验室差。一个实验室可因对数据的连续修约或测量的不灵敏等因素而造成 k 值偏小。在 k 图上按 8.3(表 6 与 7)的临界值画出的临界线,可用于考察数据的行为模式。

7.3.1.6 当按实验室分组的 k 图或 h 图表明某个实验室有好几个 k 或 h 值接近临界线时,就应考查相应的按水平分组的图。通常在按实验室分组的图中某个值看起来好象大,但实际上当在同一水平上比较,其他实验室的值与它还是很一致的。如果与其他实验室的值相差很大,就要查找原因。

7.3.1.7 除了 k 图和 h 图之外,单元平均值直方图和单元极差直方图也能揭示某些规律,例如实际存在两个不同总体。这种情况需要特殊处理,因为此处描述的方法是在总体分布是单一且是单峰的基本假定下进行的。

7.3.2 检验离群值的数值方法

7.3.2.1 对离群值的处理建议使用如下方法:

- a) 用 7.3.3 和 7.3.4 中推荐的检验判别歧离值或离群值:
 - 如果检验统计量小于或等于 5% 临界值,则接受检验的项目为正确值;
 - 如果检验统计量大于 5% 临界值,但小于或等于 1% 临界值,则称被检验的项目为歧离值,且用单星号(*)标出;
 - 如果检验统计量大于 1% 临界值,则被检验的项目称为统计离群值,且用双星号(**)标出。
- b) 调查歧离值与统计离群值是否能用某些技术错误来解释,如:
 - 测量时的失误;
 - 计算错误;
 - 登录测试结果时的简单书写错误;
 - 错误样本的分析。

当错误是属于计算或登录类型时,应用正确的值来代替可疑的结果;当错误是来自对错误样本分析时,应用正确单元的结果代替。在进行这样的更正以后,应再一次考察歧离值和离群值。如果不能用技术错误解释,从而不能对它们进行更正时,宜将这些值作为真正的离群值予以剔除,真正的离群值属于不正常的测试结果。

- c) 当歧离值和(或)统计离群值不能用技术错误解释或它们来自某个离群实验室时,歧离值仍然

作为正确项目对待而保留；而统计离群值则应被剔除，除非统计专家有充分理由决定保留它们。

- d) 按上述程序，若图 2 中的 B 表中的某个单元的数据被剔除时，则 C 表中的相应的数据也应该被剔除，反之亦然。

7.3.2.2 7.3.3 和 7.3.4 给出的检验是有两种类型的检验。柯克伦检验是对实验室内变异的检验，应该首先应用。若因此采取了任何行动，就有必要再次对剩下的数据进行检验。格拉布斯检验主要是对实验室间变异的检验，但当 n 大于 2 且柯克伦检验怀疑一个实验室内较高的变异是来自某个测试结果时，格拉布斯检验也可用来对该单元的数据进行检验。

7.3.3 柯克伦(Cochran)检验

7.3.3.1 GB/T 6379 本部分假定相对于实验室间而言，实验室内方差很小。然而经验表明情况并非总是如此，为此需对此假定的有效性进行检验。为此目的有若干检验可以使用，这里选择了柯克伦检验。

7.3.3.2 给定 p 个由相同的 n 次重复测试结果计算的标准偏差 s_i 。柯克伦检验统计量 C 定义为

$$C = \frac{s_{\max}^2}{\sum_{i=1}^p s_i^2} \dots\dots\dots (8)$$

其中 $s_{\max}$ 是这组标准差中的最大值。

- a) 如果检验统计量小于或等于 5% 临界值，则接受被检验项目为正确值；
- b) 如果检验统计量大于 5% 临界值，但小于或等于 1% 临界值，则称被检验的项目称为歧离值，且用单星号(*)标出；
- c) 如果检验统计量大于 1% 临界值，则被检验项目称为统计离群值，且用双星号(**)标出。

8.1 的表 4 给出了柯克伦检验的临界值。

柯克伦检验必须分别用于图 2 中不同水平的 C 表中。

7.3.3.3 柯克伦准则严格应用在所有标准差都是在重复性条件下，且由相同数目(n)的测试结果计算得出的情形。实际中由于数据的缺失或剔除，测试结果数可能不同。然而 GB/T 6379 本部分假定在正常组织的试验中，每个单元中测试结果数目不同所造成的影响是有限且可以忽略的，柯克伦准则中所用的 n 可取为多数单元中的测试结果数。

7.3.3.4 柯克伦准则所检验的仅对一组标准差中的最大值，从而是单侧离群值检验。当然，方差不齐也包含使某些标准差相对较小，然而小的标准差值可能很大程度受原始数据修约程度的影响，因而并不可靠。另外，似乎也没有理由拒绝一个比其他实验室精密度都要高的实验室的数据。因此柯克伦准则是合理的。

7.3.3.5 图 2 中 C 表的临界检验有时揭示一个特定实验室的标准差全部或在大多数水平下都比其他实验室的低，表明该实验室的重复性标准差要比其他实验室的低，这可能是由于它们有较好的技术或设备，也可能由于修改了或不适当地应用了标准测量方法。如果是后一种情况，应向领导小组报告，由领导小组作出决定是否应该进行更详细的调查。(在 B.1 叙述的试验中实验室 2 即是这样的例子。)

7.3.3.6 如果最大标准差经检验判为离群值，应将该值剔除而对剩下的数据再次进行柯克伦检验，此过程可以重复进行。但是当分布为近似正态的假定没有充分满足时，这样有可能导致过度的拒绝。重复应用柯克伦检验，仅在没有同时检验多个离群值的统计检验时使用。柯克伦检验不是为同时检验多个离群值而设计的，因此在下结论时要格外小心。当有两个或三个实验室的标准差都比较高，尤其是如果这是在一个水平内得出的时候，由柯克伦检验得出的结论应该仔细核查。另一方面，如果在一个实验室的不同水平下发现多个歧离值和(或)统计离群值，这表明该实验室的室内方差非常高，来自该实验室的全部数据都应该被拒绝。

7.3.4 格拉布斯(Grubbs)检验

7.3.4.1 一个离群观测值情形

给定一组数据 $x_i, i=1, 2, \dots, p$, 将其按其值大小升序排列成 $x_{(i)}$, 格拉布斯检验是检验最大观测值 $x_{(p)}$ 是否为离群值, 计算格拉布斯统计量 G_p :

$$G_p = (x_{(p)} - \bar{x})/s \dots\dots\dots (9)$$

其中

$$\bar{x} = \frac{1}{p} \sum_{i=1}^p x_{(i)} \dots\dots\dots (10)$$

$$s = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (x_{(i)} - \bar{x})^2} \dots\dots\dots (11)$$

而为检验最小观测值 $x_{(1)}$ 是否为离群值, 则计算检验统计量:

$$G_1 = (\bar{x} - x_{(1)})/s$$

- a) 如果检验统计量小于或等于 5% 临界值, 则接受被检验项目为正确值;
- b) 如果检验统计量大于 5% 临界值, 但小于或等于 1% 临界值, 则称被检验的项目称为歧离值, 且用单星号 (*) 标出;
- c) 如果检验统计量大于 1% 临界值, 则被检验项目称为统计离群值, 且用双星号 (**) 标出。

7.3.4.2 两个离群观测值情形

为检验最大的两个值是否为离群值, 计算格拉布斯检验统计量 G :

$$G = s_{p-1,p}^2/s_0^2 \dots\dots\dots (12)$$

其中:

$$s_0^2 = \sum_{i=1}^p (x_{(i)} - \bar{x})^2 \dots\dots\dots (13)$$

$$s_{p-1,p}^2 = \sum_{i=1}^{p-2} (x_{(i)} - \bar{x}_{p-1,p})^2 \dots\dots\dots (14)$$

$$\bar{x}_{p-1,p} = \frac{1}{p-2} \sum_{i=1}^{p-2} x_{(i)} \dots\dots\dots (15)$$

为检验最小的两个观测值的显著性, 计算格拉布斯检验统计量 G :

$$G = s_{1,2}^2/s_0^2 \dots\dots\dots (16)$$

其中:

$$s_{1,2}^2 = \sum_{i=3}^p (x_{(i)} - \bar{x}_{1,2})^2 \dots\dots\dots (17)$$

$$\bar{x}_{1,2} = \frac{1}{p-2} \sum_{i=3}^p x_{(i)} \dots\dots\dots (18)$$

8.2 中的表 5 给出格拉布斯检验统计量的临界值。

7.3.4.3 格拉布斯检验的应用

当分析精密度试验时, 格拉布斯检验可用于以下情形:

- a) 给定水平 j 的单元平均值见(图 2 中的 B 表)。在此情形

$$x_i = \bar{y}_{ij}$$

$$p = p_i$$

其中 j 固定。

对一个水平的数据, 对样本平均值应用 7.3.4.1 中的一个离群值情形的格拉布斯检验, 若其中最大的或最小的单元平均值经检验为离群值, 则将其剔除; 对剩下的单元平均值重复进行同样的检验。看另一个极值(若前一个检出的为最大值, 则第二次检验最小值)是否为离群值。此时不要用 7.3.4.2 中的对两个离群观测值的格拉布斯检验。当前一检验结果没有一个单元均值为离群值时, 再进行 7.3.4.2

中的对两个离群值情形的格拉布斯检验。

b) 柯克伦检验表明某个单元标准差有问题时,对该单元单个测试结果。

注:根据 7.3.2.1,如果检验统计量的值比 1% 临界值大,则称该项为统计离群值。格拉布斯检验最初用于一组单元平均值时,表 5 中的临界值是用来检验 0.5% 水平时最高的单元平均值,以及检验 0.5% 水平时最低的单元平均值。根据 7.3.2.1,这等于检验水平为 1% 时极端的单元平均值。如果发现单元平均值的极端值是统计离群值,则可将格拉布斯检验应用于其他单元平均值的极端值。也可能有人认为此时应使用单侧检验,然而 GB/T 6379 本部分所推荐的程序只使用表 5 中的临界值(显著性水平为 1% 时的双侧检验临界值),为的是使所有单元平均值都进行一致的处理。类似的也可以同样的论据来判明对歧离值的双侧 5% 临界值。

7.4 总平均值和方差的计算

7.4.1 分析方法

在 GB/T 6379 本部分中采用的分析方法包括 m 的估计以及每个不同水平下精密度的估计。计算结果对每个 j 值列在一个表中。

7.4.2 基本数据

用于计算的基本数据是图 2 所列的三个表:

——A 表为原始测试结果;

——B 表为单元平均值;

——C 表为单元离散度。

7.4.3 非空白单元

如同在 7.3.2.1 中的 d) 中叙述的那样,对于特定的水平,用于计算的非空白单元数在 B 表和 C 表中的应总是相同的。一个特殊情况是,由于数据缺失 A 表中的某个单元可能只包含一个测试结果,尽管此时 B 表不空,但 C 表中的单元则可能成为空白的,在这种情况下,可以进行以下处理:

a) 剔除这一单独测试结果,从而使 B 表和 C 表同时成为空白单元。或

b) 如果认为该数据是不应缺失的信息,就在 C 表中相应位置用一条横线(—)代替。

由于非空白单元数可随水平而不同,因此 p_j 应用下标为 j 。

7.4.4 总平均值 $\hat{m}$ 的计算

对于水平 j ,总平均值的估计为

$$\hat{m}_j = \bar{y}_j = \frac{\sum_{i=1}^p n_{ij} \bar{y}_{ij}}{\sum_{i=1}^p n_{ij}} \quad \dots\dots\dots (19)$$

7.4.5 方差的计算

对每个水平计算三个方差,即重复性方差、实验室间方差和再现性方差。

7.4.5.1 重复性方差:

$$s_{rj}^2 = \frac{\sum_{i=1}^p (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^p (n_{ij} - 1)} \quad \dots\dots\dots (20)$$

7.4.5.2 实验室间方差:

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{\bar{n}_j} \quad \dots\dots\dots (21)$$

其中:

$$\begin{aligned} s_{dj}^2 &= \frac{1}{p-1} \sum_{i=1}^p n_{ij} (\bar{y}_{ij} - \bar{y}_j)^2 \\ &= \frac{1}{p-1} \left[\sum_{i=1}^p n_{ij} (\bar{y}_{ij})^2 - (\bar{y}_j)^2 \sum_{i=1}^p n_{ij} \right] \quad \dots\dots\dots (22) \end{aligned}$$

$$\bar{n}_j = \frac{1}{p-1} \left[\sum_{i=1}^p n_{ij} - \frac{\sum_{i=1}^p n_{ij}^2}{\sum_{i=1}^p n_{ij}} \right] \dots\dots\dots (23)$$

上述计算将在附录 B.1.5 与 B.3 中用例子说明。

7.4.5.3 对于所有的 $n_{ij}=n=2$ 的特殊情况,可使用以下简单的公式:

$$s_{nj}^2 = \frac{1}{2p} \sum_{i=1}^p (y_{ij1} - y_{ij2})^2$$

$$s_{Lj}^2 = \frac{1}{p-1} \sum_{i=1}^p (\bar{y}_{ij} - \bar{y}_j)^2 - \frac{s_{nj}^2}{2}$$

这将在 B.2 中用例子说明。

7.4.5.4 由于受到误差影响,当计算结果 s_{Lj}^2 出现负值时,应将该值设置为零。

7.4.5.5 再现性方差:

$$s_{Rj}^2 = s_{nj}^2 + s_{Lj}^2 \dots\dots\dots (24)$$

7.4.6 方差依赖于 m 的情形

接下来,应该考察精密度是否依赖于 m ,若是,应确定它们之间的函数关系。

7.5 精密度值和平均水平 m 之间的函数关系的建立

7.5.1 并非总能假定在精密度与 m 之间存在某种确切的函数关系。特别是在物料的非均匀性是引起测试结果变异的不可分割的部分时,只有这种非均匀性是水平 m 的确切函数才可作上述假定。对于不同成分或来自不同生产过程的固体材料,确定的函数关系更不是一定存在的。对精密度与 m 之间是否存在某种关系必须在应用如下方法前作出决定:如果关系不存在,可对每种所考察的物料分别确定各自的精密度值。

7.5.2 7.5.3 至 7.5.9 中的推理和计算程序既能用于重复性标准差也能用于再现性标准差,为简明起见,这里仅对重复性标准差进行讨论。考虑以下 3 种类型的关系式:

- I: $s_r = bm$ (通过原点的直线);
- II: $s_r = a + bm$ (截距为正的直线);
- III: $\lg s_r = c + d \lg m$ (或 $s_r = Cm^d$), $d \leq 1$ (指数关系)。

在大多数情况下,这些关系式中至少有一个能给出满意的拟合。如果不能,进行分析的统计专家应该寻求另外的解决办法。为避免混淆,上述关系式中的常数 a, b, c, C 和 d 可用不同的下标来区别,例如对重复性记为 $a_r, b_r, \dots$; 对再现性,记为 $a_R, b_R, \dots$ 等。为简化起见,本章中这些下标均被省略。基于同样理由, s_r 也简写成 s , 而将下标记号留给水平标号 j 。

7.5.3 一般的, $d > 0$, 因而对关系式 I 与 III, 当 $m=0$ 时, 有 $s=0$ 。从试验角度看, 这似乎不能接受。因此, 当报告精密度数据时, 必须明确表示所建立的关系仅适用于实验室间精密度试验所覆盖的水平范围内。

7.5.4 当 $a=0, d=1$ 时, 上述三个关系等同, 故当 $a \approx 0, d \approx 1$ 时, 两个或所有三个拟合的关系式实际上是等价的。在此情况下, 宜采用关系式 I, 因为它允许以下简单的陈述:

“当两个测试结果的差大于 $(100 b) \%$ 时, 就应该怀疑这两个测试结果。”

用统计语言表示就是: 对所有水平, 变异系数是一个常数 $(100 s/m) \%$ 。

7.5.5 如果在 s_j 对 $\hat{m}_j$, 或 $\lg s_j$ 对 $\lg \hat{m}_j$ 的图中, 散点分布接近一条直线, 手工画出的直线即可以给出一个较为满意的解, 但最好使用数值拟合方法。对关系式 I 与 II, 建议用 7.5.6 中的方法; 而对关系式 III, 建议用 7.5.8 中的方法。

7.5.6 从统计观点来看, 拟合一条直线是很复杂的, 因为 $\hat{m}_j$ 和 s_j 都是估计值, 都有误差。但是因为斜率 b 通常很小 (小于 0.1 阶或更小), 因此 $\hat{m}$ 的误差影响很小, s 估计误差是主要的。

7.5.6.1 对回归直线参数的更好估计是使用加权回归,因为 s 的标准误差与 $s_j(s_j)$ 的预测值成比例。

加权系数应与 $1/(s_j)^2$ 成比例,其中 s_j 是水平 j 的预测的重复性标准差,然而, s_j 与需要计算的参数有关。

为使残差的加权平方和最小的估计值,严格的数学步骤相当复杂。建议使用以下已证明在实际中是满意的迭代方法。

7.5.6.2 对迭代的第 N 步,令加权系数 $W_j = 1/(s_{Nj})^2, N=0,1,2,\dots$,依次计算以下各量:

$$\begin{aligned} T_1 &= \sum_j W_j \\ T_2 &= \sum_j W_j \hat{m}_j \\ T_3 &= \sum_j W_j \hat{m}_j^2 \\ T_4 &= \sum_j W_j s_j \\ T_5 &= \sum_j W_j \hat{m}_j s_j \end{aligned}$$

对于关系式 I ($s=b\hat{m}$):

$$b = T_5/T_3$$

对于关系式 II ($s=a+b\hat{m}$):

$$a = \left[\frac{T_3 T_4 - T_2 T_5}{T_1 T_3 - T_2^2} \right] \dots\dots\dots (25)$$

$$b = \left[\frac{T_1 T_5 - T_2 T_4}{T_1 T_3 - T_2^2} \right] \dots\dots\dots (26)$$

7.5.6.3 对于关系式 I,将 $s_j=b\hat{m}_j$ 代入加权系数 $W_j=1/(s_j)^2$ 即可得到如下简单的表达式:

$$b = \frac{\sum_j (s_j/\hat{m}_j)}{q} \dots\dots\dots (27)$$

这里不需要进行迭代。

7.5.6.4 对于关系式 II,将由 7.4 中的方法得到的 s 值作为初始值 s_{0j} 。令

$$W_{0j} = 1/(s_{0j})^2 \quad (j = 1, 2, \dots, q)$$

按 7.5.6.2 中的方法计算 a_1 和 b_1 ,进而计算

$$s_{1j} = a_1 + b_1 \hat{m}_j$$

用 $W_{1j}=1/(s_{1j})^2$ 重复上述计算,可以得到

$$s_{2j} = a_2 + b_2 \hat{m}_j$$

若对加权系数 $W_{2j}=1/(s_{2j})^2$ 再次重复上述同样的步骤,所得的结果的变化将很有限。从 W_{0j} 到 W_{1j} 对消除权数的误差是很有效的,因而 s_{2j} 可看作是最终的结果。

7.5.7 $\lg s$ 的标准误差依赖于 s ,所以 $\lg s$ 关于 $\lg \hat{m}$ 的非加权回归是适宜的。

7.5.8 对于关系式 III,相应计算公式是:

$$\begin{aligned} T_1 &= \sum_j \lg \hat{m}_j \\ T_2 &= \sum_j (\lg \hat{m}_j)^2 \\ T_3 &= \sum_j \lg s_j \\ T_4 &= \sum_j (\lg \hat{m}_j)(\lg s_j) \end{aligned}$$

因此

$$c = \frac{T_2 T_3 - T_1 T_4}{q T_2 - T_1^2} \dots\dots\dots (28)$$

$$d = \frac{q T_4 - T_1 T_3}{q T_2 - T_1^2} \dots\dots\dots (29)$$

7.5.9 7.5.9.1 至 7.5.9.3 对相同的一组数据给出了拟合 7.5.2 中的关系式 I ,II 和 III 的例子。数据取自 B.3 中的案例分析,这里仅用来作为数值计算的说明。在 B.3 对此有详细的讨论。

7.5.9.1 表 1 给出拟合关系式 I 的例。

7.5.9.2 表 2 中给出拟合关系 II 式的例($\hat{m}_j, s_j$ 由 7.5.9.1 给出)。

7.5.9.3 表 3 给出了拟合关系式 III 的例。

表 1 关系式 I : $s=bm$

$\hat{m}_j$	3.94	8.28	14.18	15.59	20.41
s_j	0.092	0.179	0.127	0.337	0.393
$s_j/\hat{m}_j$	0.023 4	0.021 6	0.008 9	0.021 6	0.019 3
$b = \frac{\sum_j (s_j/\hat{m}_j)}{q}$	$\frac{0.094\ 8}{5}=0.019$				
$s=bm$	0.075	0.157	0.269	0.296	0.388

表 2 关系式 II : $s=a+bm$

W_{0j}	118	31	62	8.8	6.5
$s_1=0.058+0.009\ 0m$					
$\hat{s}_{1j}$	0.093	0.132	0.185	0.197	0.240
W_{1j}	116	57	29	26	17
$s_2=0.030+0.015\ 6m$					
$\hat{s}_{2j}$	0.092	0.159	0.251	0.273	0.348
W_{2j}	118	40	16	13	8
$s_3=0.032+0.015\ 4m$					
$\hat{s}_{3j}^{1)}$	0.093	0.160	0.251	0.273	0.348
注:加权值并非至关重要,两位有效数字已足够。					
1) 与 s_2 相差无几。					

表 3 关系式 III : $\lg s=c+d\lg m$

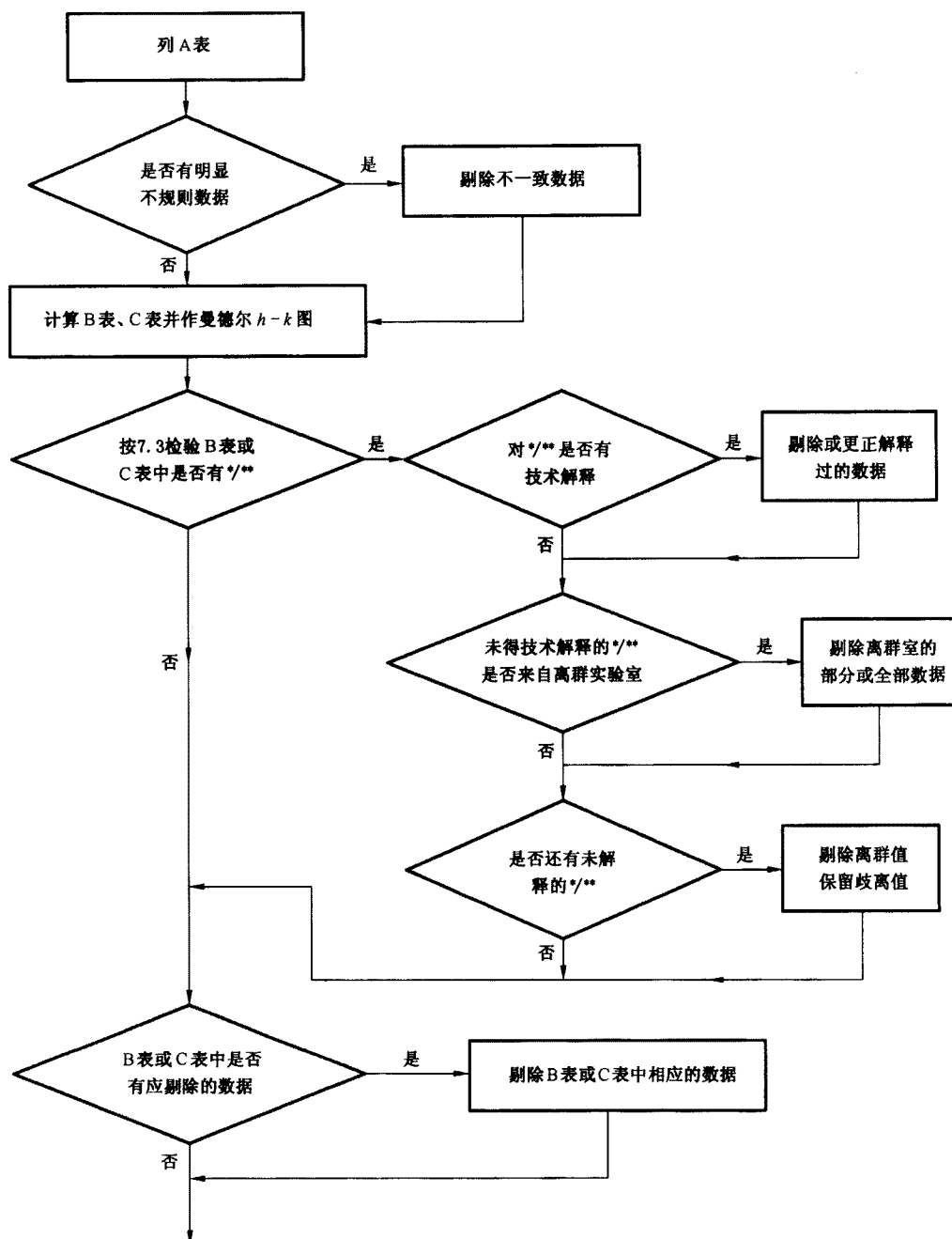
$\lg \hat{m}_j$	+0.595	+0.918	+1.152	+1.193	+1.310
$\lg s_{0j}$	-1.036	-0.747	-0.896	-0.472	-0.406
$\lg s=-1.506\ 5+0.772\lg m$ 或 $s=0.031m^{0.77}$					
s	0.089	0.158	0.239	0.257	0.316

7.6 统计分析程序的步骤

注 5: 图 3 是在 7.6 中给出的统计分析程序各步骤的流程图。

7.6.1 将所有测试结果整理成如图 2 的 A 表中(见 7.2)。建议将该表排成 p 行(标为 $i=1,2,\dots,p$, p 为参与试验的实验室数), q 列(标为 $j=1,2,\dots,q$, q 为按升序排列的水平数)。

在一致水平试验中, A 表同单元内的测试结果无需区分, 可按任意顺序排列。



备注: 图中 /* 表示歧离值和(或)离群值。

图 3 统计分析主要步骤流程图(接下页)

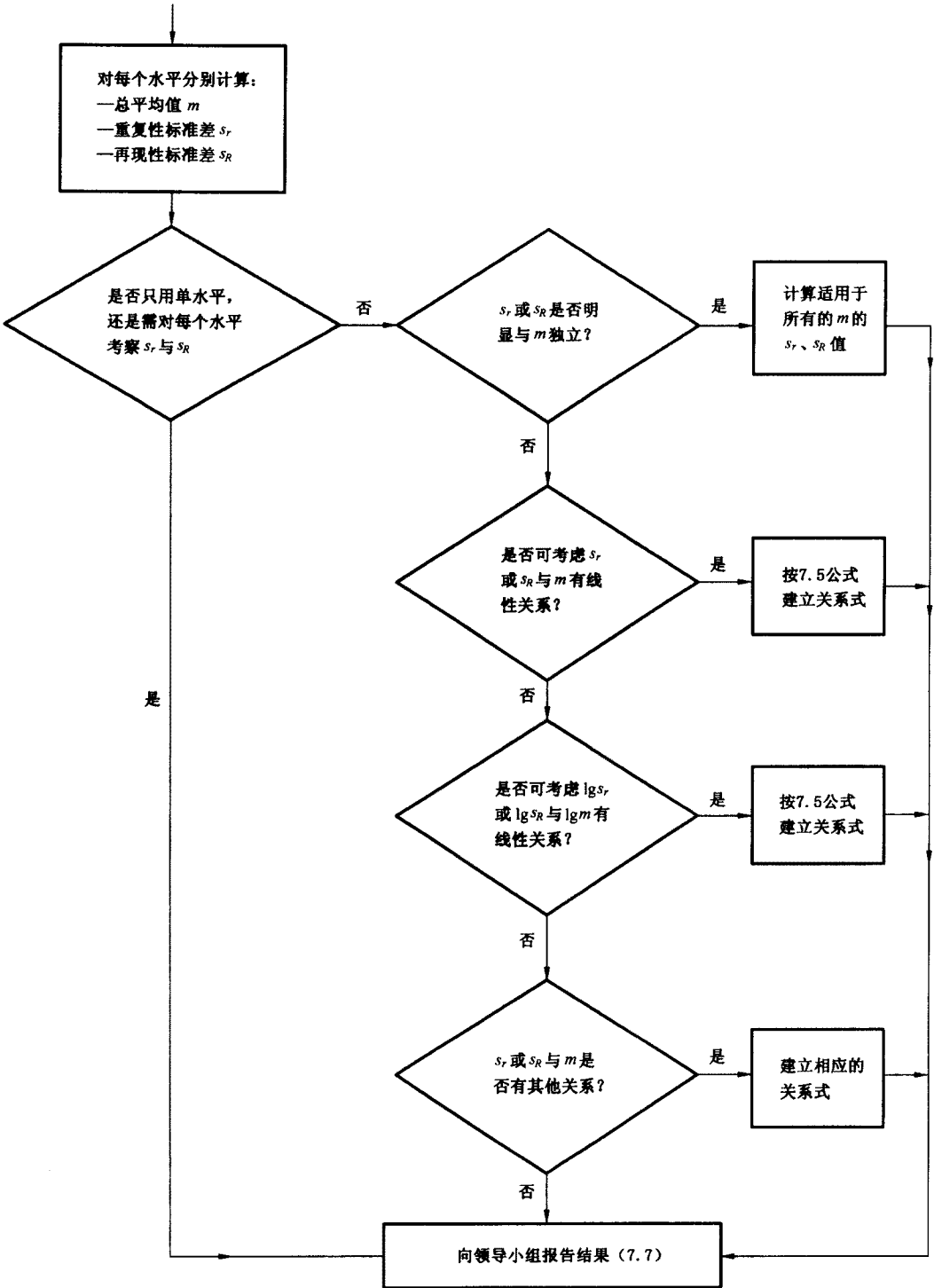


图 3(续)

7.6.2 检查 A 表中有无明显不规则的数据,剔除任何有明显错误的数据(例如在仪器测量范围之外的,或技术上不可能的数据),并向领导小组报告。有时某个特定实验室或某水平下特定单元内的测试结果明显的和其他数据不一致,对这样明显不一致的数据应该予以剔除,但也应向领导小组报告,以作进一步考虑(见 7.7.1)。

7.6.3 必要时对 A 表的数据按 7.6.2 进行更正,分别计算单元均值及单元离散度,得 B 表与 C 表。

当 A 表中的单元中只有一个测试结果时,应选用 7.4.3 中所述的其中一种方法加以处理。

7.6.4 按 7.3.1 中描述的方法绘制曼德爾 h 图和 k 图,以检验数据的一致性。这些图可以表明数据是否适宜于进一步的分析,也可以表明是否可能存在离群值和离群实验室。然而在 7.6.5 至 7.6.9 中的步骤完成以前还不能做出明确的决定。

7.6.5 对 B 表和 C 表,对每个水平,检查可能的歧离值和(或)统计离群值(见 7.3.2.1a))。对每个可疑项目用 7.3 方法进行统计检验。将每个检出的歧离值,用“*”号标出,对每个统计离群值,用“**”标出。如果没有歧离值或统计离群值,略过 7.6.6 至 7.6.10 的步骤,直接转入 7.6.11。

7.6.6 对检出的歧离值和(或)统计离群值进行检查,是否有可能对它们作技术解释。如经核实,解释令人满意,则应更正或剔除这些歧离值和(或)统计离群值,同时更正相应的表格。如果没有未得解释的歧离值和(或)统计离群值,则略过 7.6.7 至 7.6.10 的步骤,直接转入 7.6.11。

注 6:大量的歧离值和(或)统计离群值可能表示方差明显不齐或存在明显的实验室间的差异,因而怀疑该测量方法的适用性,对此也应向领导小组报告。

7.6.7 如果 B 表与 C 表中未得到解释的歧离值和(或)统计离群值的分布没有表明有离群实验室,则略过步骤 7.6.8 转入 7.6.9。

7.6.8 如果有足够的证据表明部分或全部的被怀疑的数据来自某个离群实验室,那么剔除这些数据并向给领导小组报告。

做出剔除一个特定实验室的部分或全部数据的决定,是执行统计分析任务的统计专家的职责。但该项决定应向领导小组报告以便进一步进行考虑(见 7.7.1)。

7.6.9 如果仍有尚未得到解释的或不属于某离群实验室的歧离值或离群值,则剔除统计离群值,保留歧离值。

7.6.10 如果在前面的步骤中,B 表中有数据被剔除,那么也应剔除 C 表中相应的数据;反之亦然。

7.6.11 对 B 表和 C 表中保留下来的正确值,用 7.4 中方法,对每个水平分别计算平均水平 $\hat{m}_j$ 、重复性标准差和再现性标准差。

7.6.12 如果试验是单水平的,或重复性标准差和再现性标准差需对每个水平确定而不能作为水平的函数(见 7.5.1),则略去 7.6.13 至 7.6.18 步骤直接转入 7.6.19。

注 7:以下 7.6.13 至 7.6.17 的步骤可分别应用于 s_r , s_R ,但是为简洁起见,只写成 s_r 。

7.6.13 作 s_j 对 $\hat{m}_j$ 的点图,根据图判断 s 是否依赖于 m 。如果认为 s 依赖 m ,略去步骤 7.6.14,转入 7.6.15;如果 s 不依赖于 m ,则进入 7.6.14。如果有疑问,最好两种情况都做,然后由领导小组做出决定。对此没有适用的统计检验方法,但是熟悉该测量方法的技术专家应该有足够的经验对此做出决定。

7.6.14 将 $\frac{1}{q} \sum s_j = s_r$ 作为重复性标准差的最终值,略去 7.6.15 至 7.6.18 的步骤,直接转入 7.6.19。

7.6.15 根据 7.6.13 中的图,判断 s 和 m 之间的关系是否能用直线表示,如果可以,是关系式 I ($s = bm$)还是关系式 II ($s = a + bm$)(见 7.5.2)合适?用 7.5.6 中的程序来确定参数 b 或 (a, b) 。如果线性关系令人满意,则略去步骤 7.6.16,进入 7.6.17;否则,转入 7.6.16。

7.6.16 作 $\lg s_j$ 对 $\lg \hat{m}_j$ 的点图,根据图判断 $\lg s$ 和 $\lg m$ 之间是否存在某种直线关系。如果存在,使用 7.5.8 中给出的程序拟合关系式 III ($\lg s = c + d \lg m$)。

7.6.17 如果在 7.6.15 或 7.6.16 中已建立的关系令人满意,那么对特定的 m 值, s_r (或 s_R) 最终的值得即可根据此关系式推得。然后略去 7.6.18,进入 7.6.19。

7.6.18 如果按 7.6.15 或 7.6.16 中步骤不能在 s 与 m 之间建立令人满意的关系,统计专家应决定能否建立其他类型的关系,或由于数据太不规则以至于不可能在它们之间建立关系。

7.6.19 准备一份包括基本数据和结果以及根据统计分析而得出的结论的报告,提交给领导小组。

7.3.1 中的图形对表示结果的一致性和变异性是有用的。

7.7 给领导小组的报告和领导小组做出的决定

7.7.1 统计专家的报告

完成统计分析后,统计专家应向领导小组提交一份报告。报告中应包括以下内容:

- 充分叙述从操作员和(或)测量负责人处了解到的对测量方法标准的意见;
- 充分叙述按 7.6.2 和 7.6.8 步骤被除外的离群实验室及除外的理由;
- 充分叙述所发现的每一个歧离值和(或)统计离群值,以及哪些已经得到解释、更正或删除;
- $\hat{m}_j$, s_r 和 s_R 的最终结果表及按 7.6.13、7.6.15 或 7.6.16 步骤规定所得的结论,并用在这些步骤中推荐的其中一个图说明;
- 用作统计分析的 A 表, B 表和 C 表,可作为附录。

7.7.2 领导小组采取的决定

领导小组应讨论这个报告,并对下列问题作出决定。

- 不一致的测试结果、歧离值或离群值,是否是由于对标准测量方法的不恰当的描述而引起的?
- 对被除外的离群实验室应采取什么措施?
- 离群实验室的测试结果和(或)操作员和执行负责人的意见是否能说明需要改进测量方法的标准? 如果需要,应改进哪些方面?
- 精密度试验的结果能用来确定重复性标准差和再现性标准差的值吗? 如果可以,最终值应取多少? 应该以何种的方式公布,精密度数据应用的范围是什么?

7.7.3 全面报告

执行负责人应准备一份最后经领导小组通过的包括试验工作开展理由、组织工作情况以及由统计专家提出的经同意的结论性意见。通常一些对一致性或变异性检验的图表形式也是必要的。报告应该提交给负责对此项工作进行审核和其他感兴趣的部门。

8 统计数值表

8.1 表 4 给出了柯克伦检验的临界值(见 7.3.3)。

表 4 柯克伦检验的临界值

p	$n=2$		$n=3$		$n=4$		$n=5$		$n=6$	
	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
2	—	—	0.995	0.975	0.979	0.939	0.959	0.906	0.937	0.877
3	0.993	0.967	0.942	0.871	0.883	0.798	0.834	0.746	0.793	0.707
4	0.968	0.906	0.864	0.768	0.781	0.684	0.721	0.629	0.676	0.590
5	0.928	0.841	0.788	0.684	0.696	0.598	0.633	0.544	0.588	0.506
6	0.883	0.781	0.722	0.616	0.626	0.532	0.564	0.480	0.520	0.445
7	0.838	0.727	0.664	0.561	0.568	0.480	0.508	0.431	0.466	0.397
8	0.794	0.680	0.615	0.516	0.521	0.438	0.463	0.391	0.423	0.360
9	0.754	0.638	0.573	0.478	0.481	0.403	0.425	0.358	0.387	0.329
10	0.718	0.602	0.536	0.445	0.447	0.373	0.393	0.331	0.357	0.303
11	0.684	0.570	0.504	0.417	0.418	0.348	0.366	0.308	0.332	0.281
12	0.653	0.541	0.475	0.392	0.392	0.326	0.343	0.288	0.310	0.262
13	0.624	0.515	0.450	0.371	0.369	0.307	0.322	0.271	0.291	0.243

表 4(续)

<i>p</i>	<i>n</i> =2		<i>n</i> =3		<i>n</i> =4		<i>n</i> =5		<i>n</i> =6	
	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
14	0.599	0.492	0.427	0.352	0.349	0.291	0.304	0.255	0.274	0.232
15	0.575	0.471	0.407	0.335	0.332	0.276	0.288	0.242	0.259	0.220
16	0.553	0.452	0.388	0.319	0.316	0.262	0.274	0.230	0.246	0.208
17	0.532	0.434	0.372	0.305	0.301	0.250	0.261	0.219	0.234	0.198
18	0.514	0.418	0.356	0.293	0.288	0.240	0.249	0.209	0.223	0.189
19	0.496	0.403	0.343	0.281	0.276	0.230	0.238	0.200	0.214	0.181
20	0.480	0.389	0.330	0.270	0.265	0.220	0.229	0.192	0.205	0.174
21	0.465	0.377	0.318	0.261	0.255	0.212	0.220	0.185	0.197	0.167
22	0.450	0.365	0.307	0.252	0.246	0.204	0.212	0.178	0.189	0.160
23	0.437	0.354	0.297	0.243	0.238	0.197	0.204	0.172	0.182	0.155
24	0.425	0.343	0.287	0.235	0.230	0.191	0.197	0.166	0.176	0.149
25	0.413	0.334	0.278	0.228	0.222	0.185	0.190	0.160	0.170	0.144
26	0.402	0.325	0.270	0.221	0.215	0.179	0.184	0.155	0.164	0.140
27	0.391	0.316	0.262	0.215	0.209	0.173	0.179	0.150	0.159	0.135
28	0.382	0.308	0.255	0.209	0.202	0.168	0.173	0.146	0.154	0.131
29	0.372	0.300	0.248	0.203	0.196	0.164	0.168	0.142	0.150	0.127
30	0.363	0.293	0.241	0.198	0.191	0.159	0.164	0.138	0.145	0.124
31	0.355	0.286	0.235	0.193	0.186	0.155	0.159	0.134	0.141	0.120
32	0.347	0.280	0.229	0.188	0.181	0.151	0.155	0.131	0.138	0.117
33	0.339	0.273	0.224	0.184	0.177	0.147	0.151	0.127	0.134	0.114
34	0.332	0.267	0.218	0.179	0.172	0.144	0.147	0.124	0.131	0.111
35	0.325	0.262	0.213	0.175	0.168	0.140	0.144	0.121	0.127	0.108
36	0.318	0.256	0.208	0.172	0.165	0.137	0.140	0.118	0.124	0.106
37	0.312	0.251	0.204	0.168	0.161	0.134	0.137	0.116	0.121	0.103
38	0.306	0.246	0.200	0.164	0.157	0.131	0.134	0.113	0.119	0.101
39	0.300	0.242	0.196	0.161	0.154	0.129	0.131	0.111	0.116	0.099
40	0.294	0.237	0.192	0.158	0.151	0.126	0.128	0.108	0.114	0.097

p=给定水平下的实验室数
n=每个单元的测试重复数(见 7.3.3.3)

8.2 表 5 给出了格拉布斯检验的临界值(见 7.3.4)。

表 5 格拉布斯检验的临界值

<i>p</i>	一个最大值或一个最小值		两个最大值或两个最小值	
	上 1%点	上 5%点	下 1%点	下 5%点
3	1.155	1.155	—	—
4	1.496	1.481	0.000 0	0.000 2
5	1.764	1.715	0.001 8	0.009 0
6	1.973	1.887	0.011 6	0.034 9
7	2.139	2.020	0.030 8	0.070 8
8	2.274	2.126	0.056 3	0.110 1
9	2.387	2.215	0.085 1	0.149 2
10	2.482	2.290	0.115 0	0.186 4

表 5(续)

p	一个最大值或一个最小值		两个最大值或两个最小值	
	上 1%点	上 5%点	下 1%点	下 5%点
11	2.564	2.355	0.144 8	0.221 3
12	2.636	2.412	0.173 8	0.253 7
13	2.699	2.462	0.201 6	0.283 6
14	2.755	2.507	0.228 0	0.311 2
15	2.806	2.549	0.253 0	0.336 7
16	2.852	2.585	0.276 7	0.360 3
17	2.894	2.620	0.299 0	0.382 2
18	2.932	2.651	0.320 0	0.402 5
19	2.968	2.681	0.339 8	0.421 4
20	3.001	2.709	0.358 5	0.439 1
21	3.031	2.733	0.376 1	0.455 6
22	3.060	2.758	0.392 7	0.471 1
23	3.087	2.781	0.408 5	0.485 7
24	3.112	2.802	0.423 4	0.499 4
25	3.135	2.822	0.437 6	0.512 3
26	3.157	2.841	0.451 0	0.524 5
27	3.178	2.859	0.463 8	0.536 0
28	3.199	2.876	0.475 9	0.547 0
29	3.218	2.893	0.487 5	0.557 4
30	3.236	2.908	0.498 5	0.567 2
31	3.253	2.924	0.509 1	0.576 6
32	3.270	2.938	0.519 2	0.585 6
33	3.286	2.952	0.528 8	0.594 1
34	3.301	2.965	0.538 1	0.602 3
35	3.316	2.979	0.546 9	0.610 1
36	3.330	2.991	0.555 4	0.617 5
37	3.343	3.003	0.563 6	0.624 7
38	3.356	3.014	0.571 4	0.631 6
39	3.369	3.025	0.578 9	0.638 2
40	3.381	3.036	0.586 2	0.644 5
p = 给定水平下的实验室数				
注：本表给出的是 GB/T 6379 本部分 7.3.4 中描述的格拉布斯检验方法的临界值,适用于双侧检验,其数值由附录 C 中参考文献[4]中所给的相应的单侧检验临界值推算。				

对一个离群观测值的格拉布斯检验,大于表中 1%临界值的为离群值;大于表中 5%临界值的为歧离值。

对两个离群观测值的格拉布斯检验,小于表中 1%临界值的为离群值;小于表中 5%临界值的为歧离值。

8.3 表 6 和表 7 给出了曼德尔的 h 和 k 统计量的临界值。

表 6 显著性水平为 1% 时曼德尔 h 和 k 统计量的临界值

p	h	k								
		n								
		2	3	4	5	6	7	8	9	10
3	1.15	1.71	1.64	1.58	1.53	1.49	1.46	1.43	1.41	1.39
4	1.49	1.91	1.77	1.67	1.60	1.55	1.51	1.48	1.45	1.43
5	1.72	2.05	1.85	1.73	1.65	1.59	1.55	1.51	1.48	1.46
6	1.87	2.14	1.90	1.77	1.68	1.62	1.57	1.53	1.50	1.47
7	1.98	2.20	1.94	1.79	1.70	1.63	1.58	1.54	1.51	1.48
8	2.06	2.25	1.97	1.81	1.71	1.65	1.59	1.55	1.52	1.49
9	2.13	2.29	1.99	1.82	1.73	1.66	1.60	1.56	1.53	1.50
10	2.18	2.32	2.00	1.84	1.74	1.66	1.61	1.57	1.53	1.50
11	2.22	2.34	2.01	1.85	1.74	1.67	1.62	1.57	1.54	1.51
12	2.25	2.36	2.02	1.85	1.75	1.68	1.62	1.58	1.54	1.51
13	2.27	2.38	2.03	1.86	1.76	1.68	1.63	1.58	1.55	1.52
14	2.30	2.39	2.04	1.87	1.76	1.69	1.63	1.58	1.55	1.52
15	2.32	2.41	2.05	1.87	1.76	1.69	1.63	1.59	1.55	1.52
16	2.33	2.42	2.05	1.88	1.77	1.69	1.63	1.59	1.55	1.52
17	2.35	2.44	2.06	1.88	1.77	1.69	1.64	1.59	1.55	1.52
18	2.36	2.44	2.06	1.88	1.77	1.70	1.64	1.59	1.56	1.52
19	2.37	2.44	2.07	1.89	1.78	1.70	1.64	1.59	1.56	1.53
20	2.39	2.45	2.07	1.89	1.78	1.70	1.64	1.60	1.56	1.53
21	2.39	2.46	2.07	1.89	1.78	1.70	1.64	1.60	1.56	1.53
22	2.40	2.46	2.08	1.90	1.78	1.70	1.65	1.60	1.56	1.53
23	2.41	2.47	2.08	1.90	1.78	1.71	1.65	1.60	1.56	1.53
24	2.42	2.47	2.08	1.90	1.79	1.71	1.65	1.60	1.56	1.53
25	2.42	2.47	2.08	1.90	1.79	1.71	1.65	1.60	1.56	1.53
26	2.43	2.48	2.09	1.90	1.79	1.71	1.65	1.60	1.56	1.53
27	2.44	2.48	2.09	1.90	1.79	1.71	1.65	1.60	1.56	1.53
28	2.44	2.49	2.09	1.91	1.79	1.71	1.65	1.60	1.57	1.53
29	2.45	2.49	2.09	1.91	1.79	1.71	1.65	1.60	1.57	1.53
30	2.45	2.49	2.10	1.91	1.79	1.71	1.65	1.61	1.57	1.53
p = 给定水平下的实验室数										
n = 给定水平下每个实验室的测试重复数										

表 7 显著性水平为 5% 时曼德尔 h 和 k 统计量的临界值

p	h	k								
		n								
		2	3	4	5	6	7	8	9	10
3	1.15	1.65	1.53	1.45	1.40	1.37	1.34	1.32	1.30	1.29
4	1.42	1.76	1.59	1.50	1.44	1.40	1.37	1.35	1.33	1.31
5	1.57	1.81	1.62	1.53	1.46	1.42	1.39	1.36	1.34	1.32
6	1.66	1.85	1.64	1.54	1.48	1.43	1.40	1.37	1.35	1.33
7	1.71	1.87	1.66	1.55	1.49	1.44	1.41	1.38	1.36	1.34

表 7(续)

p	h	k								
		n								
		2	3	4	5	6	7	8	9	10
8	1.75	1.88	1.67	1.56	1.50	1.45	1.41	1.38	1.36	1.34
9	1.78	1.90	1.68	1.57	1.50	1.45	1.42	1.39	1.36	1.35
10	1.80	1.90	1.68	1.57	1.50	1.46	1.42	1.39	1.37	1.35
11	1.82	1.91	1.69	1.58	1.51	1.46	1.42	1.39	1.37	1.35
12	1.83	1.92	1.69	1.58	1.51	1.46	1.42	1.40	1.37	1.35
13	1.84	1.92	1.69	1.58	1.51	1.46	1.43	1.40	1.37	1.35
14	1.85	1.92	1.70	1.59	1.52	1.47	1.43	1.40	1.37	1.35
15	1.86	1.93	1.70	1.59	1.52	1.47	1.43	1.40	1.38	1.36
16	1.86	1.93	1.70	1.59	1.52	1.47	1.43	1.40	1.38	1.36
17	1.87	1.93	1.70	1.59	1.52	1.47	1.43	1.40	1.38	1.36
18	1.88	1.93	1.71	1.59	1.52	1.47	1.43	1.40	1.38	1.36
19	1.88	1.93	1.71	1.59	1.52	1.47	1.43	1.40	1.38	1.36
20	1.89	1.94	1.71	1.59	1.52	1.47	1.43	1.40	1.38	1.36
21	1.89	1.94	1.71	1.60	1.52	1.47	1.44	1.41	1.38	1.36
22	1.89	1.94	1.71	1.60	1.52	1.47	1.44	1.41	1.38	1.36
23	1.90	1.94	1.71	1.60	1.53	1.47	1.44	1.41	1.38	1.36
24	1.90	1.94	1.71	1.60	1.53	1.48	1.44	1.41	1.38	1.38
25	1.90	1.94	1.71	1.60	1.53	1.48	1.44	1.41	1.38	1.36
26	1.90	1.94	1.71	1.60	1.53	1.48	1.44	1.41	1.38	1.36
27	1.91	1.94	1.71	1.60	1.53	1.48	1.44	1.41	1.38	1.36
28	1.91	1.94	1.71	1.60	1.53	1.48	1.44	1.41	1.38	1.36
29	1.91	1.94	1.72	1.60	1.53	1.48	1.44	1.41	1.38	1.36
30	1.91	1.94	1.72	1.60	1.53	1.48	1.44	1.41	1.38	1.36

p = 给定水平下的实验室数
 n = 给定水平下每个实验室的测试重复数

附录 A

(规范性附录)

GB/T 6379 所用的符号与缩略语

- a 关系式 $s=a+bm$ 中的截距
- A 用来计算估计值的不确定度的系数
- b 关系式 $s=a+bm$ 中的斜率
- B 表示一个实验室测试结果与总平均值的偏差分量(偏倚的实验室分量)
- B_0 表示在中间精密度条件下所有因素皆保持不变时 B 的分量
- $B_{(1)}, B_{(2)}, \dots$ 表示在中间精密度条件下, 因素发生改变时 B 的分量
- c 关系式 $\lg s=c+d\lg m$ 中的截距
- C, C', C'' 检验统计量
- $C_{crit}, C'_{crit}, C''_{crit}$ 用于统计检验的临界值
- CD_p 概率 P 的临界差
- CR_p 概率 P 的临界极差
- d 关系式 $\lg s=c+d\lg m$ 中的斜率
- e 发生在每次测试结果中随机误差分量
- f 临界极差系数
- $F_p(\nu_1, \nu_2)$ 自由度为 ν_1 和 ν_2 的 F 分布的 p 分位数
- G 格拉布斯检验统计量
- h 曼得尔实验室间一致性检验统计量
- k 曼得尔实验室内一致性检验统计量
- LCL 控制下限(行动限或警戒限)
- m 测试特性的总平均值; 水平
- M 在中间精密度条件中考虑的因素数
- N 交互作用数
- n 一个实验室在一个水平(即一个单元中)上的测试结果数
- p 参加实验室间试验的实验室数
- P 概率
- q 在实验室间试验中测试特性的水平数
- r 重复性限
- R 再现性限
- RM 标准物料
- s 标准差的估计值
- $\hat{s}$ 标准差的预测值
- T 总和
- t 测试目标个数或组数
- UCL 控制上限(行动限或警戒限)
- W 加权回归中的权数
- w 一组测试结果的极差
- x 用于格拉布斯检验的数据
- y 测试结果

- $\bar{y}$ 测试结果的算术平均值
 - $\bar{\bar{y}}$ 测试结果的总平均值
 - α 显著性水平
 - β 第二类错误概率
 - γ 再现性标准差与重复性标准差的比值(σ_R/σ_r)
 - Δ 实验室偏倚
 - $\hat{\Delta}$ Δ 的估计值
 - δ 测量方法偏倚
 - $\hat{\delta}$ δ 的估计值
 - λ 两个实验室偏倚或两个测量方法偏倚之间的可检出的差
 - μ 测试特性的真值或接受参照值
 - ν 自由度
 - ρ 方法 A 和方法 B 的重复性标准差之间的可检出的比
 - σ 标准差的真值
 - τ 表示从上次校准始由时间变化引起的测试结果变异的分量
 - ϕ 方法 A 和方法 B 的实验室间均方的平方根可检出的比
 - $\chi_p^2(\nu)$ 自由度为 ν 的 χ^2 分布的 p 分位数
- 用作下标的符号:
- C 校准差
 - E 设备差
 - i 实验室标识
 - $I()$ 精密度的中间度量;括号内表示中间情形类型
 - j 水平的标识(GB/T 6379.2);测试或因素的标识(GB/T 6379.3)
 - k 实验室 i , 水平为 j 的测试结果的标识
 - L 实验室间
 - m 可检出偏倚的标识
 - M 试样间
 - O 操作员差
 - P 概率
 - r 重复性
 - R 再现性
 - T 时间差
 - W 实验室内
- 1, 2, 3, ... 测试结果按获得顺序的编号
 - (1), (2), (3), ... 测试结果按数值大小递增顺序的编号

附 录 B
(资料性附录)
精密度试验统计分析的实例

B.1 例 1:煤中硫含量的确定(多水平,不包含缺失值与离群值数据)

B.1.1 背景

a) 测量方法

确定煤中的硫含量,试验结果以质量百分数表示。

b) 资料来源

Tomkins, S. S.《化学工业工程》(参见附录 C 中[6])

c) 简述

8 个实验室参与试验,根据引用的资料来源中所描述的标准化的测量方法对测试结果进行分析。对每个水平实验室 1 报告了 4 个试验结果,实验室 5 报告了 4~5 个测试结果,其余的实验室都报告 3 个测试结果。

d) 图形表示

一般应将曼德尔 h 和 k 统计量绘制成图形,但在本例中,这些图形不是很重要,故为节省篇幅,将其省略。B.3 中的实例对曼德尔点图进行了全面的说明和讨论。

B.1.2 原始数据

原始数据用图 2 中 A 表的格式(参见 7.2.8)列于表 B.1 中,以质量百分数[%(m/m)]表示,没有任何一个数据有特殊的标记。

这些数据的图形在图 B.1 到 B.4 中给出。

表 B.1 原始数据:煤中硫的含量

实验室 i	水 平 j			
	1	2	3	4
1	0.71	1.20	1.68	3.26
	0.71	1.18	1.70	3.26
	0.70	1.23	1.68	3.20
	0.71	1.21	1.69	3.24
2	0.69	1.22	1.64	3.20
	0.67	1.21	1.64	3.20
	0.68	1.22	1.65	3.20
3	0.66	1.28	1.61	3.37
	0.65	1.31	1.61	3.36
	0.69	1.30	1.62	3.38
4	0.67	1.23	1.68	3.16
	0.65	1.18	1.66	3.22
	0.66	1.20	1.66	3.23

表 B. 1(续)

实验室 <i>i</i>	水 平 <i>j</i>			
	1	2	3	4
5	0.70	1.31	1.64	3.20
	0.69	1.22	1.67	3.19
	0.66	1.22	1.60	3.18
	0.71	1.24	1.66	3.27
	0.69	—	1.68	3.24
6	0.73	1.39	1.70	3.27
	0.74	1.36	1.73	3.31
	0.73	1.37	1.73	3.29
7	0.71	1.20	1.69	3.27
	0.71	1.26	1.70	3.24
	0.69	1.26	1.68	3.23
8	0.70	1.24	1.67	3.25
	0.65	1.22	1.68	3.26
	0.68	1.30	1.67	3.26

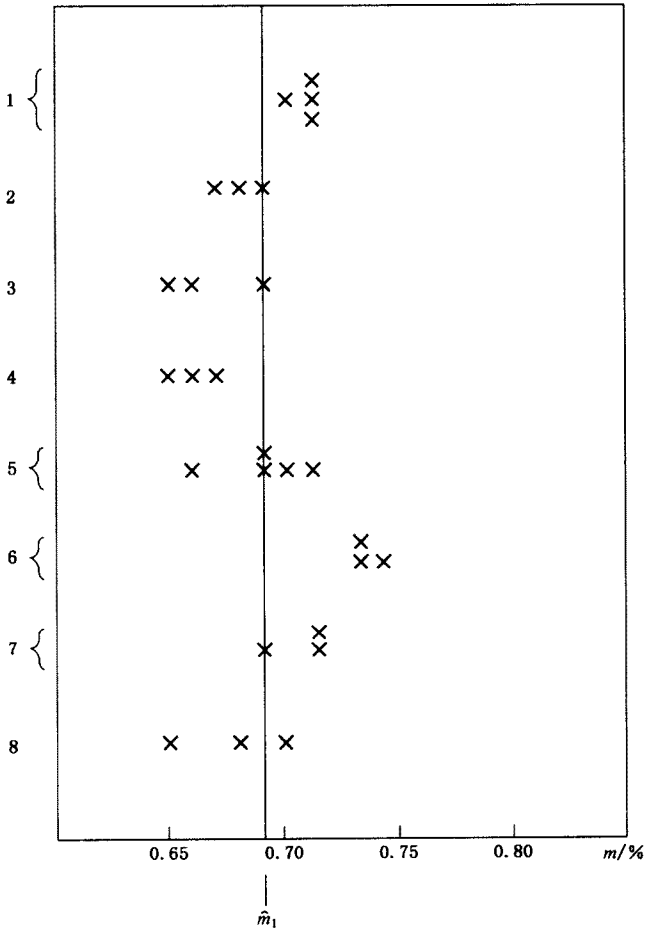


图 B. 1 样本 1, 煤中硫的含量

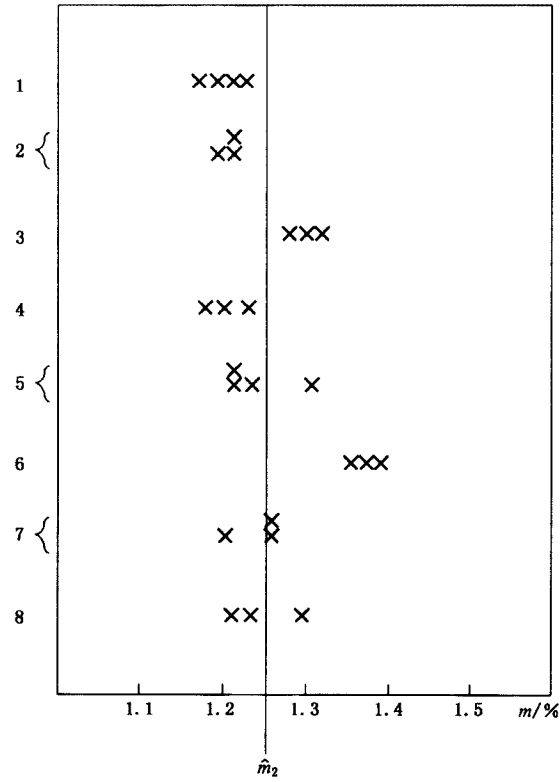


图 B.2 样本 2,煤中硫的含量

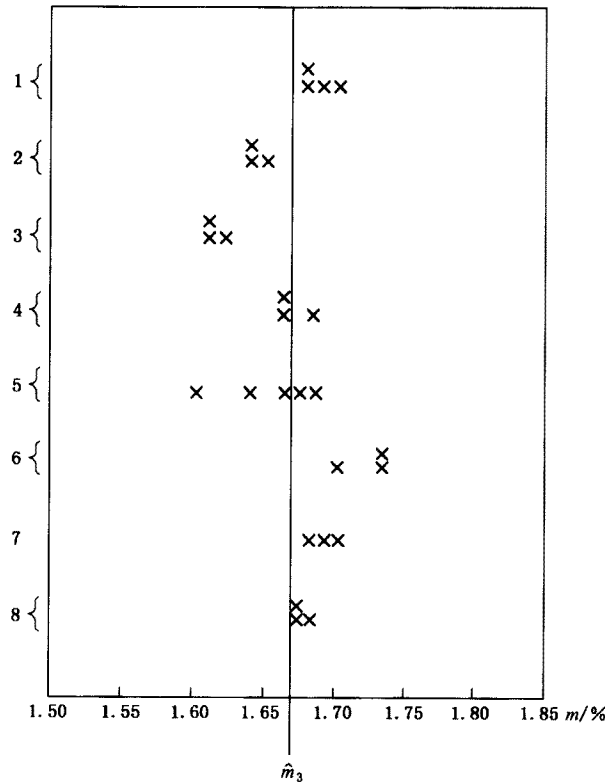


图 B.3 样本 3,煤中硫的含量

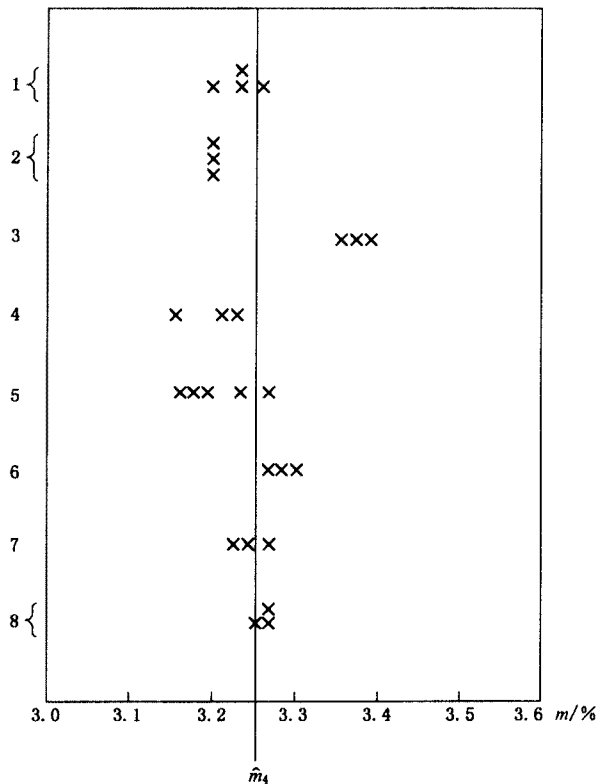


图 B.4 样本 4,煤中硫的含量

注 8: 表 B.1 中所引用的试验,没有指明实验室应该做多少次试验,只给出了应测试的最小次数。根据 GB/T 6379 本部分所推荐程序,对实验室 1 和 5,宜从给定的测试值中随机抽取 3 个数据,使所有单元中的测试结果刚好等于 3。然而,为说明计算程序也能适用于那些测试结果数可能不等的情形,本例中保留了所有的测试结果。读者可以将每个单元的测试结果用随机抽取的办法减少到 3 个,经过计算,可以验证对 $\hat{m}_i$, s_r 和 s_R 的值影响甚微。

B.1.3 单元平均值 y_{ij} 的计算

单元平均值按(7.2.9)图 2 中 B 表的格式列于表 B.2 中,单位为质量百分数[%(m/m)]。

表 B.2 煤中硫含量的单元平均值

实验室 i	水 平 j							
	1		2		3		4	
	$\bar{y}_{ij}$	n_{ij}	$\bar{y}_{ij}$	n_{ij}	$\bar{y}_{ij}$	n_{ij}	$\bar{y}_{ij}$	n_{ij}
1	0.708	4	1.205	4	1.688	4	3.240	4
2	0.680	3	1.217	3	1.643	3	3.200	3
3	0.667	3	1.297	3	1.613	3	3.370	3
4	0.660	3	1.203	3	1.667	3	3.203	3
5	0.690	5	1.248	4	1.650	5	3.216	5
6	0.733	3	1.373	3	1.720	3	3.290	3
7	0.703	3	1.240	3	1.690	3	3.247	3
8	0.677	3	1.253	3	1.673	3	3.257	3

B.1.4 标准差 s_{ij} 的计算

标准差按(7.2.10)图 2 中 C 表的格式列于表 B.3 中,单位为质量百分数[%(m/m)]。

表 B.3 煤中硫含量的标准差

实验室 <i>i</i>	水 平 <i>j</i>							
	1		2		3		4	
	<i>s_{ij}</i>	<i>n_{ij}</i>	<i>s_{ij}</i>	<i>n_{ij}</i>	<i>s_{ij}</i>	<i>n_{ij}</i>	<i>s_{ij}</i>	<i>n_{ij}</i>
1	0.005	4	0.021	4	0.010	4	0.028	4
2	0.010	3	0.006	3	0.006	3	0.000	3
3	0.021	3	0.015	3	0.006	3	0.010	3
4	0.010	3	0.025	3	0.012	3	0.038	3
5	0.019	5	0.043	4	0.032	5	0.038	5
6	0.006	3	0.015	3	0.017	3	0.020	3
7	0.012	3	0.035	3	0.010	3	0.021	3
8	0.025	3	0.042	3	0.006	3	0.006	3

B.1.5 一致性和离群值的检查

$n=3, p=8$, 柯克伦检验 5% 临界值为 0.516; 1% 临界值为 0.615。

对水平 1, 实验室 8 的 s 最大:

$$\sum s^2 = 0.001\ 82; \text{检验统计量值} = 0.347$$

对水平 2, 实验室 5 的 s 最大:

$$\sum s^2 = 0.006\ 36; \text{检验统计量值} = 0.287$$

对水平 3, 实验室 5 的 s 最大:

$$\sum s^2 = 0.001\ 72; \text{检验统计量值} = 0.598$$

对水平 4, 实验室 4 的 s 最大:

$$\sum s^2 = 0.004\ 63; \text{检验统计量值} = 0.310$$

这表明水平 3 的一个单元可以看作是歧离值, 没有离群值。该歧离值仍然参与后续计算。

将格拉布斯检验应用于单元平均值, 表 B.4 中给出了这些平均值。没有单个的歧离值或离群值, 在水平 2 和 4 中, 对实验室 3 和 6, 根据双高检验, 最高测试结果为歧离值。这些歧离值在分析中予以保留。

表 B.4 对单元平均值的格拉布斯检验

水 平	单个低值	单个高值	两个低值	两个高值	检验类型
1	1.24	1.80	0.539	0.298	格拉布斯检验 统计量
2	0.91	2.09	0.699	0.108	
3	1.67	1.58	0.378	0.459	
4	0.94	2.09	0.679	0.132	
歧离值	2.126	2.126	0.110 1	0.110 1	格拉布斯检验 临界值
	2.274	2.274	0.056 3	0.056 3	

B.1.6 $\hat{m}_j, s_{ij}$ 和 s_{Rj} 的计算

以水平 1 为例, 7.4.4 和 7.4.5 中所定义的诸方差计算如下。

实验室数 $p=8$

$$T_1 = \sum n_i \bar{y}_i = 18.642$$

$$T_2 = \sum n_i (\bar{y})^2 = 12.883\ 7$$

$$T_3 = \sum n_i = 27$$

$$T_4 = \sum n_i^2 = 95$$

$$T_5 = \sum (n_i - 1) s_i^2 = 0.004\ 411$$

$$s_r^2 = \frac{T_5}{T_3 - p} = 0.000\ 232\ 2$$
$$s_L^2 = \left[\frac{T_2 T_3 - T_1^2}{T_3(p-1)} - s_r^2 \right] \left[\frac{T_3(p-1)}{T_3^2 - T_4} \right] = 0.000\ 460\ 3$$
$$s_R^2 = s_L^2 + s_r^2 = 0.000\ 692\ 5$$
$$\hat{m} = \frac{T_1}{T_3} = 0.690\ 44$$
$$s_r = 0.015\ 24$$
$$s_R = 0.026\ 32$$

类似的可对水平 2,3 和 4 进行计算,其结果列于表 B. 5。

B. 1. 7 精密度与 *m* 的关系

对表 B. 5 中的数据进行检查,没有显示出它们与 *m* 有任何依赖关系。因而可用它们的平均值。

表 B. 5 煤中硫含量的 $\hat{m}_j, s_{rj}$ 和 s_{Rj} 值

水平 <i>j</i>	<i>p_j</i>	$\hat{m}_j$	<i>s_{rj}</i>	<i>s_{Rj}</i>
1	8	0.690	0.015	0.026
2	8	1.252	0.029	0.061
3	8	1.667	0.017	0.035
4	8	3.250	0.026	0.058

B. 1. 8 结论

测量方法精密度(以质量的百分数表示)可引述如下:

重复性标准差:*s_r* = 0.022;

再现性标准差:*s_R* = 0.045。

这些值的适当范围为 0.69% ~ 3.25%(*m/m*)。这些值是通过有 8 个实验室参与的一致水平试验获得的,所测硫含量的值在上述范围内,试验中共检测到 4 个歧离值,但在分析中予以保留。

B. 2 例 2: 沥青的软化点(多水平,包含缺失值数据)

B. 2. 1 背景

a) 测量方法

使用环和球来确定沥青的软化点。

b) 资料来源

测试焦油及其产品的标准方法:沥青部分;使用中性甘油的方法 PT3(参见附录 C 中的参考文献[5])。

c) 物料

根据参考文献[5]中“样本”一章中的有关沥青部分所规定的程序,从收集到的多批商业沥青中对物料进行选择和准备。

d) 简述

需要测量的特性是温度,以摄氏度表示。有 16 个实验室参与协同试验。分别在 87.5℃, 92.5℃, 97.5℃ 和 102.5℃ 测量 4 个样品,这包含了通常的商业产品范围。但是水平 2 选择了错误的物料,它的平均温度为 96℃,与水平 3 的相似。实验室 5 在水平 2 最初对样本使用了错误的方法(所测试的第一个值),它没有足够的物料进行重测。实验室 8 没有水平 1 的样本(水平 4 有 2 个样本)。

e) 图形表示

一般应将曼德尔 *h* 和 *k* 统计量绘制成图,但在本例中,限于篇幅再次被省略。B. 3 中的实例将对曼德尔的点图进行了全面的说明和讨论。

B.2.2 原始数据

这些数据用图 2 中 A 表的格式(参见 7.2.8)列于表 B.6 中,单位为摄氏度。

表 B.6 原始数据:沥青的软化点(℃)

实验室 <i>i</i>	水 平 <i>j</i>			
	1	2	3	4
1	91.0	97.0	96.5	104.0
	89.6	97.2	97.0	104.0
2	89.7	98.5	97.2	102.6
	89.8	97.2	97.0	103.6
3	88.0	97.8	94.2	103.0
	87.5	94.5	95.8	99.5
4	89.2	96.8	96.0	102.5
	88.5	97.5	98.0	103.5
5	89.0	97.2	98.2	101.0
	90.0	—	98.5	100.2
6	88.5	97.8	99.5	102.2
	90.5	97.2	103.2	102.0
7	88.9	96.6	98.2	102.8
	88.2	97.5	99.0	102.2
8	—	96.0	98.4	102.6
	—	97.5	97.4	103.9
9	90.1	95.5	98.2	102.8
	88.4	96.8	96.7	102.0
10	86.0	95.2	94.8	99.8
	85.8	95.0	93.0	100.8
11	87.6	93.2	93.6	98.2
	84.4	93.4	93.9	97.8
12	88.2	95.8	95.8	101.7
	87.4	95.4	95.4	101.2
13	91.0	98.2	98.0	104.5
	90.4	99.5	97.0	105.6
14	87.5	97.0	97.1	105.2
	87.8	95.5	96.6	101.8
15	87.5	95.0	97.8	101.5
	87.6	95.2	99.2	100.9
16	88.8	95.0	97.2	99.5
	85.0	93.2	97.8	99.8
注:没有明显的歧离值和统计离群值。				

B.2.3 单元平均值

这些数据按(7.2.9)图 2 中 B 表的格式列于 B.7 中,单位为摄氏度。

图 B.5 中给出了这些数据的图示。

表 B.7 沥青的软化点的单元平均值 (°C)

实验室 <i>i</i>	水 平 <i>j</i>			
	1	2	3	4
1	90.30	97.10	96.75	104.00
2	89.75	97.85	97.10	103.10
3	87.75	96.15	95.00	101.25
4	88.85	97.15	97.00	103.00
5	89.50	—	98.35	100.60
6	89.50	97.50	101.35	102.10
7	88.55	97.05	98.60	102.50
8	—	96.75	97.90	103.25
9	89.25	96.15	97.45	102.40
10	85.90	95.10	93.90	100.30
11	86.00	93.30	93.75	98.00
12	87.80	95.60	95.60	101.45
13	90.70	98.85	97.50	105.05
14	87.65	96.25	96.85	103.50
15	87.55	95.10	98.50	101.20
16	86.90	94.10	97.50	99.65

注：已剔除了 $i=5, j=2$ 这一单元的测试结果(见 7.4.3)。

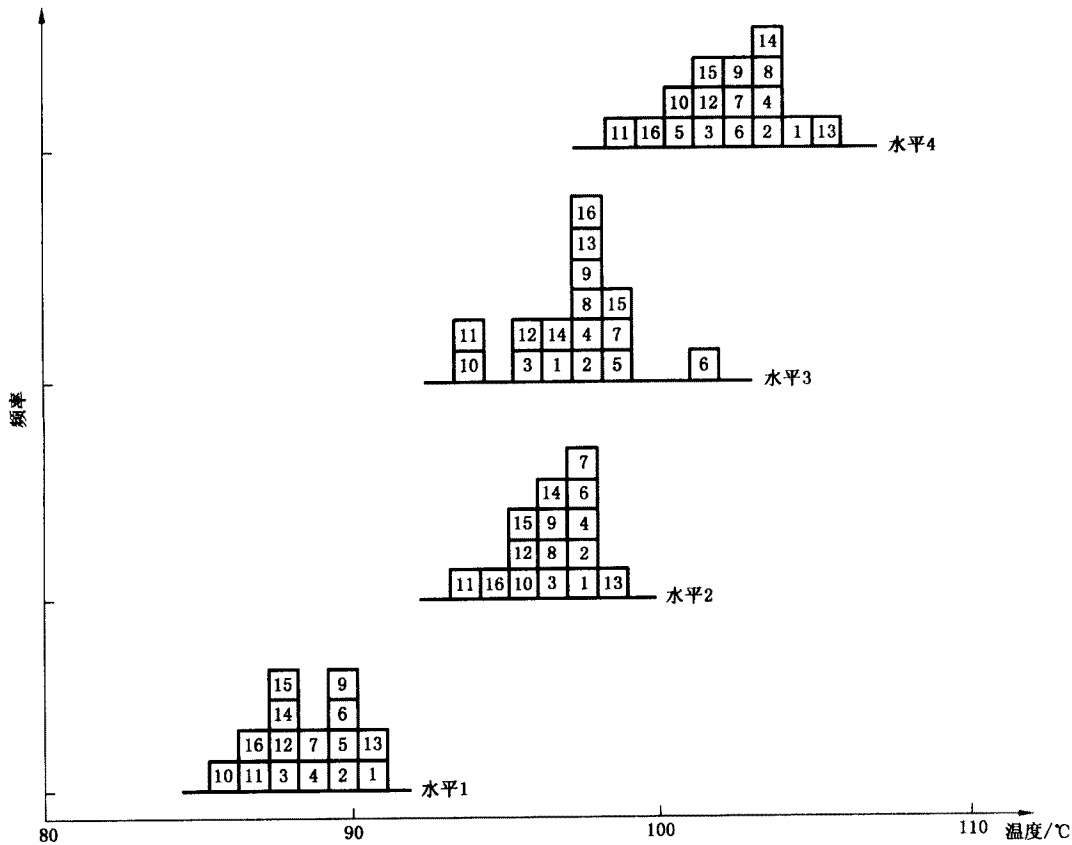


图 B.5 沥青软化点的单元平均值

B.2.4 单元内绝对差

在本例中，每个单元有两个测试结果，故可用绝对差用来表示变异。单元的绝对差按图 2 中 C 表的格式列于表 B.8，单位为摄氏度(见 7.2.10)。

图 B.6 中给出了这些数据的图示。

表 B.8 沥青软化点的单元内绝对差

实验室 <i>i</i>	水 平 <i>j</i>			
	1	2	3	4
1	1.4	0.2	0.5	0.0
2	0.1	1.3	0.2	1.0
3	0.5	3.3	1.6	3.5
4	0.7	0.7	2.0	1.0
5	1.0	—	0.3	0.8
6	2.0	0.6	3.7	0.2
7	0.7	0.9	0.8	0.6
8	—	1.5	1.0	1.3
9	1.7	1.3	1.5	0.8
10	0.2	0.2	1.8	1.0
11	3.2	0.2	0.3	0.4
12	0.8	0.4	0.4	0.5
13	0.6	1.3	1.0	1.1
14	0.3	1.5	0.5	3.4
15	0.1	0.2	1.4	0.6
16	3.8	1.8	0.6	0.3

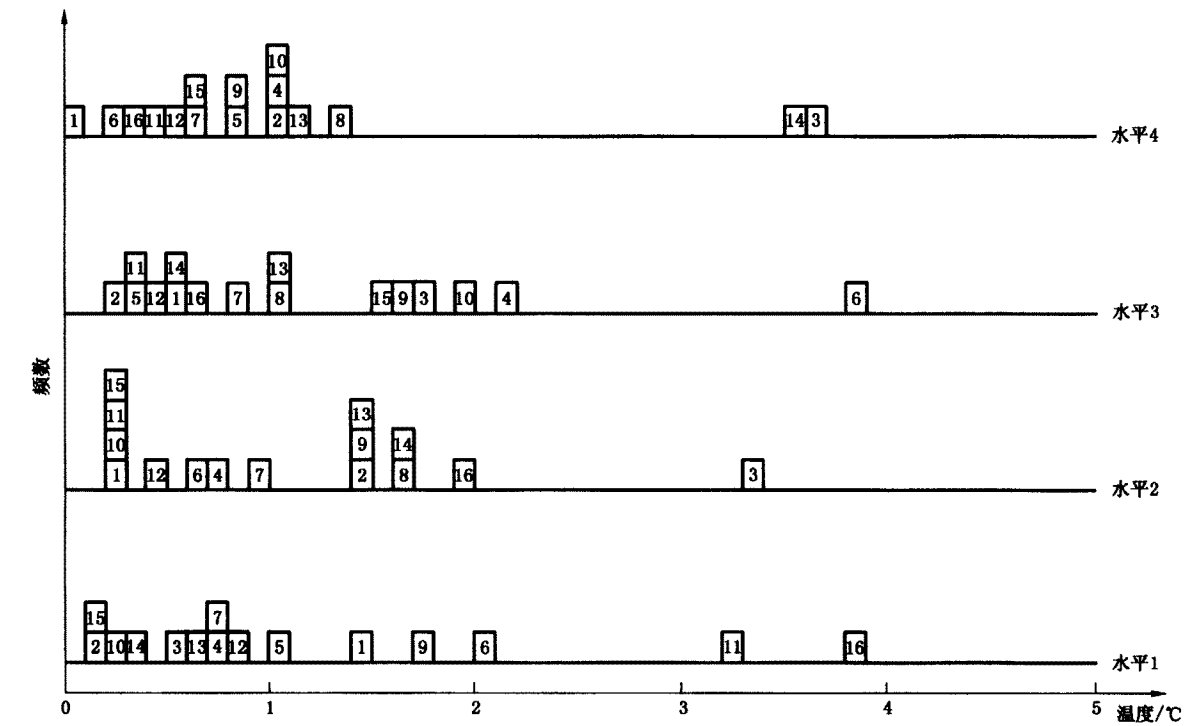


图 B.6 沥青软化点的单元内绝对差

B.2.5 对一致性和离群值的检查

应用柯克伦检验,计算得到的检验统计量 C 的值,列于表 B.9。

对 $n=2, p=15$, 显著性水平为 5% 时的临界值(参看 8.1)为 0.471, $p=16$ 时为 0.452, 表明没有歧离值。

将格拉布斯检验用于单元平均值,也没有发现有单个或成对歧离值或离群值存在。

表 B.9 柯克伦检验统计量 C 的值

水平 j	1	2	3	4
C	0.391(15)	0.424(15)	0.434(16)	0.380(16)
注: 括号内给出的是实验室数。				

表 B.10 对单元平均值的格拉布斯检验

水平; n	单个低值	单个高值	两个低值	两个高值	检验类型
1;15	1.69	1.56	0.546	0.662	格拉布斯检验 统计量
2;15	2.04	1.77	0.478	0.646	
3;16	1.76	2.27	0.548	0.566	
4;16	2.22	1.74	0.500	0.672	
歧离值					格拉布斯检验 临界值
$n=15$	2.549	2.549	0.336 7	0.336 7	
$n=16$	2.585	2.585	0.360 3	0.360 3	
离群值					
$n=15$	2.806	2.806	0.253 0	0.253 0	
$n=16$	2.852	2.852	0.276 7	0.276 7	

B.2.6 $\hat{m}_j, s_{rj}$ 和 s_{Rj} 的计算

这些值按 7.4.4 和 7.4.5 中的方法计算。

以水平 1 为例,具体计算如下。为了简化计算,所有数据都减去 80.00。以下方法适用于每个单元重复 $n=2$ 次的情形。

实验室数 $p=15$, 重复次数 $n=2$

$$\begin{aligned}T_1 &= \sum \bar{y}_i = 125.950\ 0 \\T_2 &= \sum (\bar{y}_i)^2 = 1\ 087.977\ 5 \\T_3 &= \sum (y_{i1} - y_{i2})^2 = 36.910\ 0 \\s_r^2 &= \frac{T_3}{2p} = 1.230\ 3 \\s_L^2 &= \left[\frac{pT_2 - T_1^2}{p(p-1)} \right] - \frac{s_r^2}{2} = 1.557\ 5 \\s_R^2 &= s_L^2 + s_r^2 = 2.787\ 8 \\\hat{m} &= \frac{T_1}{p} (+80.00) = 88.396\ 6 \\s_r &= 1.109\ 2 \\s_R &= 1.669\ 7\end{aligned}$$

所有 4 个水平的计算结果列于表 B.11。

表 B.11 沥青软化点 $\hat{m}_j$, s_{rj} 和 s_{Rj} 值的计算

水平 j	p_j	$\hat{m}_j$ (°C)	s_{rj}	s_{Rj}
1	15	88.40	1.109	1.670
2	15	96.27	0.925	1.597
3	16	97.07	0.993	2.010
4	16	101.96	1.004	1.915

B.2.7 精密度与 m 的关系

对表 B.11 进行粗略的考察揭示不出任何两者之间有明显的依赖关系,或许再现性标准差稍有一些迹象。考虑到 m 值的范围太小,不足以表示任何显著的依赖关系。从该测量的本质看,它们与 m 的依赖关系也不会存在。因此可以得出下面的结论,在正常商业材料范围内精密度不依赖于 m ,因而这些量的平均值可以作为重复性和再现性标准差的最终值。

B.2.8 结论

就实际应用而言,测量方法的精密度值不依赖于材料的水平,相应的数值为:

重复性标准差: $s_r = 1.0^\circ\text{C}$

再现性标准差: $s_R = 1.8^\circ\text{C}$

B.3 例 3: 木馏油的热滴定(多水平,包含离群值数据)

B.3.1 背景

a) 资料来源

测试沥青及其产品的标准方法;木馏油部分;方法 Co.18(参见附录 C 中的参考文献^[5])。

b) 物料

根据参考文献^[6]中“样本”一章中的木馏油部分所规定的程序,从收集到的多批商业木馏油中对物料进行选择 and 准备。

c) 简述

热滴定法在化学分析中是一种标准的测量方法,测试结果用质量的百分数表示。有 9 个实验室参加对 5 个样品每个进行 2 次重复测量。选择的测量样本覆盖一般商业应用中的的正常范围,即 4,8,12,16 和 20[%(m/m)]水平内。实际中通常测试结果只保留到小数点后一位,但是在这个试验中,要求操作员将测量结果保留到小数点后两位。

B.3.2 原始数据

这些数据用图 2 中 A 表的格式(参见 7.2.8)表示。列于表 B.12 中,单位为质量百分数。

实验室 1 的测试结果总是比其他实验室的高,并且在某些水平下还高出许多。

实验室 6 在水平 5 下的第二个测试结果值得怀疑,该记录值似乎更符合在水平 4 下的结果。

这些问题将在 B.3.5 中进一步讨论。

表 B.12 原始数据:木馏油的热滴定

实验室 i	水 平 j									
	1		2		3		4		5	
1	4.44	4.39	9.34	9.34	17.40	16.90	19.23	19.23	24.28	24.00
2	4.03	4.23	8.42	8.33	14.42	14.50	16.06	16.22	20.40	19.91
3	3.70	3.70	7.60	7.40	13.60	13.60	14.50	15.10	19.30	19.70
4	4.10	4.10	8.93	8.80	14.60	14.20	15.60	15.50	20.30	20.30
5	3.97	4.04	7.89	8.12	13.73	13.92	15.54	15.78	20.53	20.88

表 B. 12(续)

实验室 <i>i</i>	水 平 <i>j</i>									
	1		2		3		4		5	
6	3.75	4.03	8.76	9.24	13.90	14.06	16.42	16.58	18.56	16.58
7	3.70	3.80	8.00	8.30	14.10	14.20	14.90	16.00	19.70	20.50
8	3.91	3.90	8.04	8.07	14.84	14.84	15.41	15.22	21.10	20.78
9	4.02	4.07	8.44	8.17	14.24	14.10	15.14	15.44	20.71	21.66

B. 3.3 单元平均值

这些值按(7.2.9)图 2 中 B 表的格式列于表 B. 13 中,单位为质量百分数。

表 B. 13 木馏油热滴定的单元平均值

实验室 <i>i</i>	水 平 <i>j</i>				
	1	2	3	4	5
1	4.415	9.340	17.150**	19.230**	24.140*
2	4.130	8.375	14.460	16.140	20.155
3	3.700	7.500	13.600	14.800	19.500
4	4.100	8.865	14.400	15.550	20.300
5	4.005	8.005	13.825	15.660	20.705
6	3.890	9.000	13.980	16.500	17.570
7	3.750	8.150	14.150	15.450	20.100
8	3.905	8.055	14.840	15.315	20.940
9	4.045	8.305	14.170	15.290	21.185
* 看作歧离值。					
** 看作统计离群值。					

B. 3.4 单元内绝对差

这些值记为 w_{ij} 按(7.2.10)图 2 中 C 表的格式列于表 B. 14 中,单位为质量百分数。

表 B. 14 木馏油热滴定的单元极差

实验室 <i>i</i>	水 平 <i>j</i>				
	1	2	3	4	5
1	0.05	0.00	0.50	0.00	0.28
2	0.20	0.09	0.08	0.16	0.49
3	0.00	0.20	0.00	0.60	0.40
4	0.00	0.13	0.40	0.10	0.00
5	0.07	0.23	0.19	0.24	0.35
6	0.28	0.48	0.16	0.16	1.98*
7	0.10	0.30	0.10	1.10*	0.80
8	0.01	0.03	0.00	0.19	0.32
9	0.05	0.27	0.14	0.30	0.95
* 看作歧离值。					

B. 3.5 一致性和离群性的检查

在图 B. 7 和图 B. 8 中给出了计算得到的曼德尔 h 和 k 一致统计量(参见 7.3.1)的值,水平线表示与 8.3 中的曼德尔临界值相对应的临界线。

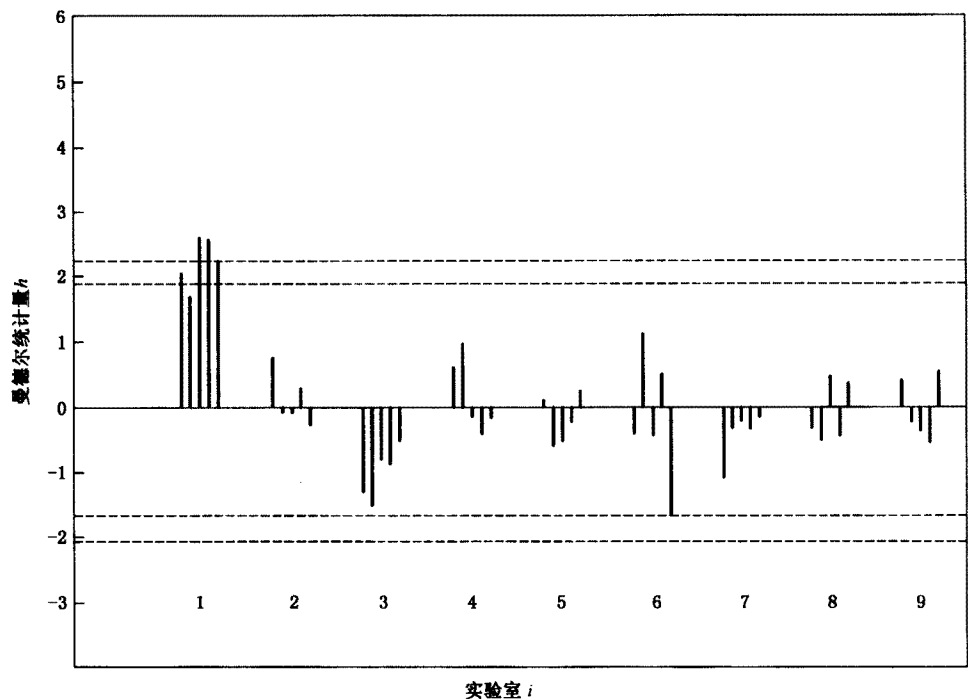


图 B.7 木馏油滴定:按实验室进行分组的实验室间一致性曼德尔统计量 h

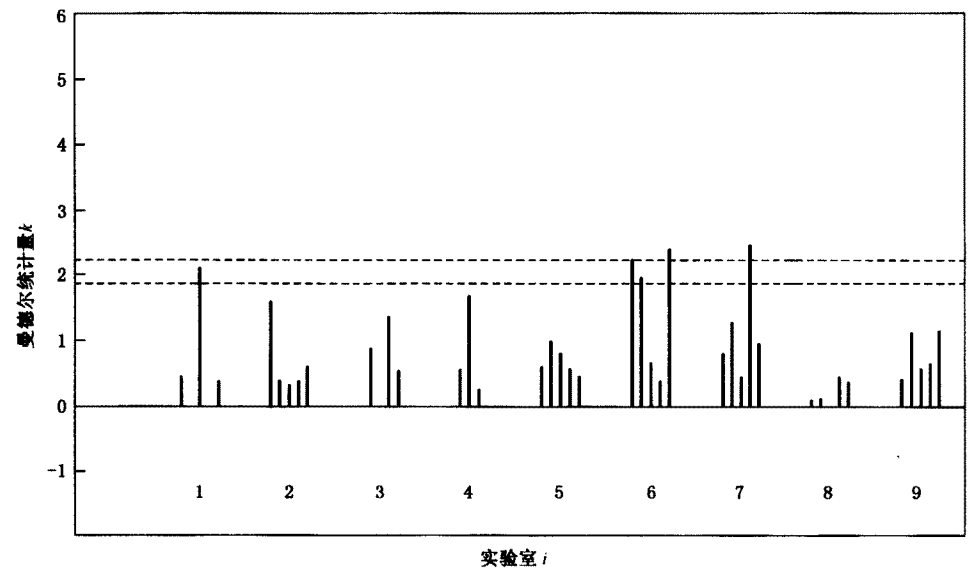


图 B.8 木馏油滴定:按实验室进行分组的实验室间一致性曼德尔统计量 k

h 图(图 B.7)清晰地表示出实验室 1 在各水平下的试验结果要比其他实验室相应水平的试验结果高。这样的结果要求管理此项实验室间研究的委员会给予关注。如果对这些测试结果找不到合理的解释,那么委员会成员应依据其他的或非统计的考虑进行判断,来决定在计算精密密度值时是保留还是剔除该实验室。

k 图(图 B.8)显示了在实验室 6 和 7 中的两次重复测试结果之间有相当大的变异。然而,这些测试结果似乎还没有严重到要求采取特别行动的程度,只是要求找出可能的解释,对这些试验结果采取必要的补救措施。

应用柯克伦检验得到下面结果。

在水平 4,根据绝对差 1.10,检验统计量值为 $1.10^2/1.814\ 9=0.667$ 。

在水平 6,根据绝对差 1.98,检验统计量值为 $1.98^2/6.166\ 3=0.636$ 。

对于 $p=9$,柯克伦检验的临界值在显著水平为 5%时是 0.638,1%时是 0.754。

在水平 4 时,1.10 显然是一个歧离值;水平 5 下的 1.98 也很接近 5%的临界值,所以有可能也是一个歧离值。由于这两个值与所有其他值相差甚大,并且它们也使得用在柯克伦检验统计量中的分母增大,所以它们都被看作是歧离值而标以“*”号。然而到目前为止,还没有足够的证据拒绝这两个值,尽管曼德尔 k 图(图 B.8)也对这两个值提出了质疑。

将格拉布斯检验应用于单元平均值,给出表 B.15 的结果。

表 B.15 对单元平均值的格拉布斯检验

水 平	单个低值	单个高值	两个低值	两个高值	检验类型
1	1.36	1.95	0.502	0.356	格拉布斯检验 统计量
2	1.57	1.64	0.540	0.395	
3	0.86	2.50	—	—	
4	0.91	2.47	—	—	
5	1.70	2.10	0.501	0.318	
歧离值	2.215	2.215	0.149 2	0.149 2	格拉布斯检验 临界值
离群值	2.387	2.387	0.085 1	0.085 1	

对水平 3 和 4,因为一个观测值的格拉布斯检验表明有离群值,所以就不需要做两个观测值的格拉布斯检验(参见 7.3.4)。

实验室 1 对水平 3 和 4 的单元平均值值是离群值,该实验室在水平 5 时单元平均值值也很大,这在曼德尔 h 图中也可以明显的看出(图 B.7)。

进一步的调查得知,实验室 6 在水平 5 至少有一个样本误用了水平 4 的样本。由于该单元的绝对差也值得怀疑,因此决定剔除这一对试验结果。由于剔除了实验室 6 的水平 5 的这对值,实验室 1 水平 5 的测试结果现在显得更加可疑。

因为这些测试结果,也因为不能肯定测量的是什么样的物料,所以决定拒绝这对来自于实验室 6 水平 5 的测试结果,同时也拒绝来自于离群实验室 1 的所有结果。将这些测试结果排除以后,把 8 个实验室水平 4 的柯克伦检验统计量与临界值(5%时是 0.680)进行比较,就不再认为是歧离值,予以保留。

B.3.6 $\hat{m}_j, s_{rj}$ 和 s_{Rj} 的计算

$\hat{m}_j, s_{rj}$ 和 s_{Rj} 的计算值在表 B.16 中给出,单位为质量百分数。计算使用 7.4.4 和 7.4.5 中的程序,这些值在计算过程中剔除了实验室 1 的测试结果以及实验室 6 水平 5 的一对测试结果。

表 B.16 木馏油热滴定的 $\hat{m}_j, s_{rj}$ 和 s_{Rj} 计算值

水平 j	p_j	$\hat{m}_j$	s_{rj}	s_{Rj}
1	8	3.94	0.092	0.171
2	8	8.28	0.179	0.498
3	8	14.18	0.127	0.400
4	8	15.59	0.337	0.579
5	7	20.41	0.393	0.637

B.3.7 精密度与 m 的关系

根据表 B.16,随着 m 值的加大显然标准差也在增加,所以可能建立它们之间的某种关系。熟悉测量方法的化学家也认为精密度很可能是依赖水平 m 的。

拟合函数关系的实际计算这里并没有给出,因为它已经在 7.5.9 中计算 s_r 时详细地进行了阐述。 s_{rj} 和 s_{Rj} 与 $\hat{m}_j$ 的值画在图 B.9 中。

根据图 B.9,显然水平 3 的值非常分散,不能通过任何程序得到改进(参看 7.5.2)。

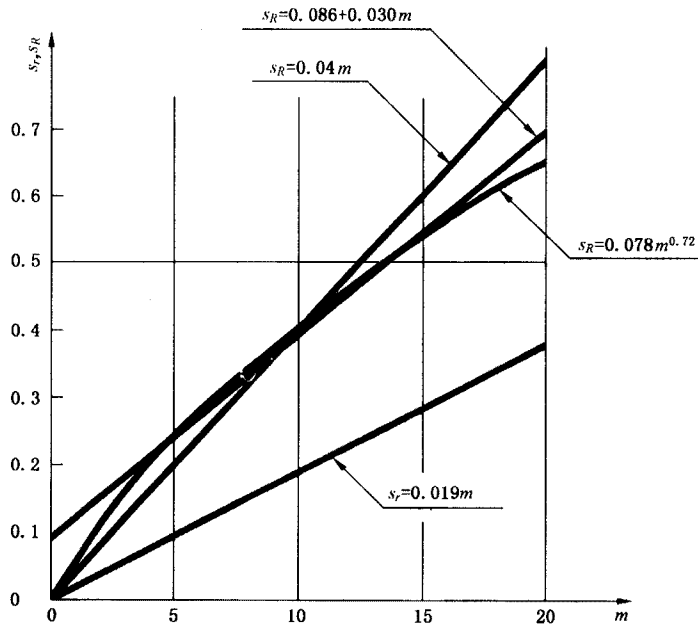


图 B.9 表 B.16 中数据的 s_{rj} 和 s_{Rj} 与 $\hat{m}_j$ 的图,表示 7.5 中数据拟合的关系

对于重复性而言,直线通过原点似乎是合适的。

对于再现性而言,所有的三条直线拟合的都很好。其中关系Ⅲ的拟合效果最好。熟悉木馏油标准测试方法的人员能在其中选择一种最适宜的关系式。

B.3.8 精密度的最终值

经过适当的舍入,精密度的最终值为

重复性标准差, $s_r = 0.019m$

再现性标准差用以下关系式计算:

$$s_R = 0.086 + 0.030m$$

或

$$s_R = 0.078m^{0.72}$$

B.3.9 结论

在统计上没有什么理由偏好 B.3.8 中的 s_R 的两个等式中的任一个,应该由领导小组决定使用哪一个。

应该调查实验室 1 的离群试验结果的原因。

似乎这个精密度实验不令人非常满意。9 个实验室中有一个作为离群实验室被排除,另外一个测试了一个错误的样本。水平 3 的物料似乎也被选错,它几乎和水平 4 的结果完全相同而不是介于水平 2 和 4 之间。此外,水平 3 的物料似乎性质上稍稍有些不同,或许比其他物料质地更均匀。这个试验值得重复进行,只是对不同的水平需要更仔细地选择物料。

附 录 C
(资料性附录)
参 考 文 献

- [1] ISO Guide 33:1989, *Use of certified reference materials*.
- [2] ISO Guide 35:1989, *Certification of reference materials—General and statistical principles*.
- [3] ASTM E691-87, *Standard practice for Conducting an interlaboratory Study to Determine the Precision of a Test Method*. American Society for Testing and Materials, Philadelphia, PA, USA.
- [4] GRUBBS, F. E. and BECK, G. Extension of sample sizes and percentage points for significance test of outlying observations. *Technometrics*, 14, 1972, pp. 847-854.
- [5] “Standard Methods for Testing Tar and its Products”. 7th Ed. Standardisation of Tar Products Tests Committee, 1979.
- [6] YOMKING, S. S. *Industrial and Engineering Chemistry* (Analytical edition), 14, 1942, pp. 141-145.
- [7] GRUBBS, F. E. Procedures for detecting outlying observation in samples. *Technometrics*, 11, 1969, pp. 1-21.
- [8] ISO 3534-2:1993, 统计学 词汇和符号 第2部分:统计质量控制
Statistics—Vocabulary and symbols—Part 2: Statistical quality control.
- [9] ISO 3534-3:1985, 统计学 词汇和符号 第3部分:实验设计
Statistics—Vocabulary and symbols—Part 3: Design of experiments.
- [10] ISO 5725-3:1994, 测量方法与结果的准确度(正确度与精确度) 第3部分:标准测量方法精密度的中间度量
Accuracy (trueness and precision) of measurement methods and results—Part 3: Intermediate measures of the precision of a standard measurement method.
- [11] ISO 5725-4:1994, 测量方法与结果的准确度(正确度与精确度) 第4部分:确定标准测量方法正确度的基本方法
Accuracy (trueness and precision) of measurement methods results—Part 4: Basic methods for the determination of the trueness of a standard and measurement method.
- [12] ISO 5725-5:1998, 测量方法与结果的准确度(正确度与精确度) 第5部分:确定标准测量方法正确度的可替代方法
Accuracy (trueness and precision) of measurement methods and results—Part 5: Alternative methods for the determination of the precision of a standard measurement method.
- [13] ISO 5725-6:1994, 测量方法与结果的准确度(正确度与精确度) 第6部分:准确度值的实际应用
Accuracy (trueness and precision) of measurement methods results—Part 6: Use in practice of accuracy values.