

doi: 10.7690/bgzdh.2022.09.009

一种指控装备试验数据的检验与统计分析框架

姜伟杰, 权冀川, 刘勇, 罗晨

(陆军工程大学指挥控制工程学院, 南京 210007)

摘要: 为进一步提高指控装备试验数据使用的有效性和工作效率, 基于当前指控装备效能评估的军事需求及试验数据分析现状, 提出一种试验数据的检验与统计分析框架。该框架在分析指控装备试验数据类型的基础上, 根据样本量大小, 确定合适的试验数据异常值检验及统计方法; 其中, 对于小样本试验数据采用基于 Bootstrap 的统计分析方法。结果表明, 该框架可为下一阶段指控装备效能评估的数据使用提供处理方法指导。

关键词: 指控装备; 试验数据; 异常值检验; 统计分析; Bootstrap 方法

中图分类号: TJ06 **文献标志码:** A

A Framework for Test and Statistical Analysis of Test Data of Command and Control Equipment

Jiang Weijie, Quan Jichuan, Liu Yong, Luo Chen

(Command and Control Engineering College, Army Engineering University of PLA, Nanjing 210007, China)

Abstract: In order to further improve the utilization effectiveness and efficiency of the test data of command and control equipment, based on the current military requirements of the effectiveness evaluation of command and control equipment and the current situation of test data analysis, a test and statistical analysis framework of test data is proposed. Based on the analysis of the test data types of command and control equipment, the framework determines the appropriate test and statistical methods of outliers in test data according to the sample size. Bootstrap-based data statistical analysis method is used for small sample test data. The results show that the framework can provide the guidance of data processing method for the effectiveness evaluation of command and control equipment in the next stage.

Keywords: command and control equipment; test data; outlier test; statistical analysis; Bootstrap method

0 引言

指控装备既是指挥信息系统的核心组成部分, 又是军队武器装备的重要组成部分, 作为战斗力的“倍增器”和作战体系的“神经中枢”, 在信息化战争中发挥着关键性作用。为全面了解指控装备的性能和作战效能, 必须进行装备试验。基于试验数据对装备的性能和效能进行评估, 是我军装备论证、装备建设和装备保障过程中不可或缺的环节。指控装备试验需要投入的经济成本高、资源耗费大、实施周期长。只有充分利用好各种试验数据, 才能有效提高指控装备试验的有效性和工作效率。同时, 在指控装备效能评估过程中, 存在试验次数不够、数据采集不完备、异常数据干扰等因素, 导致指控装备试验数据样本量小、数据不完整、利用率低等现实问题, 增加了指控装备效能定量评估的难度; 因此, 需要结合指控装备试验数据的内容特点, 提出切实可行的解决方法, 以提高试验数据利用率, 发挥试验数据的应有作用。

1 装备试验数据分析的现状

随着新军事变革的不断深入, 军队深化体制改革工作扎实推进, 装备发展速度不断提升, 一批功能先进的主战武器装备相继列装部队^[1]。指控装备近几年更新换代速度非常快, 装备试验需求量剧增, 在效能评估指标体系、效能评估理论方法等方面的研究成果越来越多; 但绝大多数文献更多聚焦于理论层面的研究或翻译外军的成果, 对于试验实施过程中存在的许多困难和问题, 不少研究成果只是笼统地讨论, 实际上是脱离试验本身的, 所以很难用于指导实践^[2]。现代装备一般体系结构复杂、造价高昂, 试验的难度大、费用高, 而复杂装备常采用小批量研制模式, 获得的历史数据和有参考价值的信息较少; 因此, 要获得大量的试验数据样本非常困难, 普遍存在小样本的问题, 传统的大样本统计方法不再适用于该类数据的分析^[3-4]。

针对经典测试性评估方法存在的所需故障样本量大、评估精度低及测试性虚拟仿真实验数据可信

收稿日期: 2022-05-24; 修回日期: 2022-06-23

作者简介: 姜伟杰(1991—), 男, 江苏人, 硕士, 从事效能评估研究。E-mail: jaywage@163.com。

度不高等缺点，王璇等^[5]提出一种基于信息融合的装备测试性评估方法，将虚拟仿真实验与实物试验结合起来，增大了用于测试性评估的信息量，提高了评估的精度，同时减小了评估风险。由于试验成本和试验时间等因素的影响，有时只能获得极少的试验样本；李文丽等^[6]指出可对极小子样试验数据采用虚拟增广法，利用 Bootstrap 方法得到增广再生子样，利用数据分析的方法对增广再生子样进行可靠性评估；叶伟等^[7]提出一种基于极大熵准则和虚拟样本的思想，利用相关的历史数据、专家经验和实际小样本数据等先验信息，建立由虚拟正品制造系统与虚拟次品生成系统组成的复杂装备极大熵计算机虚拟研制系统，然后基于费希尔信息量建立虚拟总体的检验模型，优化筛选出最佳的虚拟总体，从而解决因样本少导致的数据少、因信息匮乏导致的可靠性预测与质量控制难的问题；张春丽等^[8]在典型数据生成过程中结合蒙特卡洛模拟对文献中常见的 Bootstrap 检验流程的检验功效、检验水平等问题展开比较研究，提出应用 Bootstrap 检验流程的若干建议。针对试验数据中存在的异常值问题，刘金娣等^[9]围绕成年人工效学用户测试数据中的 7 种异常值情形，比较了异常值的检验方法，得出当数据差异较小时，格鲁布斯检验法和 *t* 检验法检验效果较好；而当数据差异较大时，跳跃度检验法的检验效果更好。针对航天装备试验鉴定数据量大、种类多、增长快、处理速度快、蕴藏价值大等特点，李亚楠等^[10]提出了大数据管理与分析平台技术实现的方法，通过深入展现数据中潜在的或复杂的模式和关系，实现数据价值利用的最大化。

综合来看，装备试验中普遍存在小样本的情况，大多数文献重点研究了小样本条件下的试验数据分析问题，对小样本数据本身可能存在的异常值情况重视不足，而异常值数据很容易影响统计分析结果。同时，也存在不同种类数据大样本、小样本并存的实际情况。笔者针对指控装备试验数据的特点，提出一种对不同样本量数据进行检验和统计分析的框架。

2 指控装备试验数据检验与统计分析框架

对指控装备进行效能评估时，装备的试验数据是评估指标计算的原始数据或要素数据。为在试验数据与评估指标之间建立有效的关联和映射关系，对应到不同类别评估指标的计算。同时，为便于指控装备试验数据的分类组织和维护管理，可将指控

装备试验数据分为下述类型：时间类数据，比率、概率类数据，误码率、丢包率类数据，数量类数据，能力、质量、满意度类数据，速率类数据，最值类数据和直接观测类数据，如表 1 所示。

表 1 指控装备试验数据的类型

序号	数据类型	数据范围
1	时间类数据	计算 2 个时刻的差值，即得到要测量的试验数据。试验时，要测量多次，得到多个数值，再计算它们的平均值。把得到的平均值作为试验数据的统计结果。在此过程中，记录下所有数据的最大值、最小值及中位数、分位数等数据，作为最大或最小类时间数据
2	比率、概率类数据	向指控装备输入一定数量（如 <i>n</i> 个）指令或信息，待其处理完毕后，收集处理结果，记录符合评估指标要求结果的数量（设为 <i>m</i> 个）。计算 <i>m/n</i> 所得的百分比作为数据测量值
3	误码率、丢包率类数据	该类试验数据往往可以使用专用的测量或测试设备直接进行测量，得到的测量结果即为需要的试验数据，不必再进行额外的计算处理。而且，这类试验数据也作为某些装备性能指标的标称值出现
4	数量类数据	指控装备试验时，需要通过统计、计数得到的试验数据。有时，这类试验数据作为比率、概率类数据的输入数据
5	能力、质量、满意度类数据	这类数据属于定性类数据，一般不能通过指控装备试验直接得到。测量时需要领域专家的参与或干预，一般采用专家打分（百分制、或针对评价集打评语）方式得到试验数据
6	速率类数据	这类数据一般可以通过一段时间的试验得到。在一段时间内，持续地给指控装备输入要处理的信息，时间结束时，统计指控装备实际处理的信息量，求其相对于时间的平均值即得到试验数据。有的试验数据也可以直接从产品说明书中得到
7	最值类数据	这类试验数据一般要通过压力测试的方式进行测量，即在很短时间内向指控装备输入大量的待处理信息，统计其最多能够处理的信息数量。有些指标也可以直接从产品说明书中得到
8	直接观测类数据	这类数据可以在指控装备运行时从显示终端或系统状态记录中直接观测到。其结果往往是以文字记录、等级评价等形式出现。直接观测类数据必须由人参与才能得到试验结果。而且，这类数据往往与能力、质量、满意度类等定性数据有较强的关联关系，可以作为确定这些定性数据的依据

而指控装备效能评估需要的指标数据，一般来自上述指控装备试验数据的统计分析结果。按照传统的数理统计理论，要获得这些参数的分布及置信区间，需要依赖足够数量的大样本；但是，指控装备试验的实际情况往往是仅有有限的历史记录，或者缺乏符合实际情况的试验环境导致部分类型的数据难以获得充分的数据量。针对不同的试验数据样本量大小的情况，只有采取合适的方法进行数据处理，才能减少误差。试验数据检验与统计分析框架如图 1 所示。

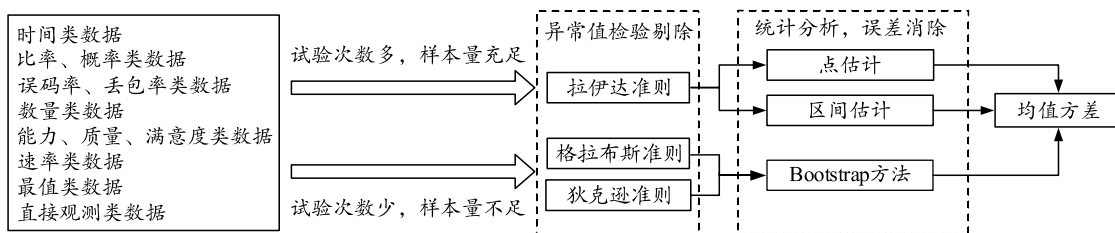


图 1 试验数据检验与统计分析框架

3 大本量试验数据的异常值检验与统计

3.1 大样本数据异常值检验

在指控装备试验中，时间类数据、速率类数据和误码率、丢包率类数据一般可以在同一次试验过程中，通过延长试验的时间、提高测量的次数或调整输入变量的数值获得较多的测量数据，能够形成大样本数据。在装备试验领域，根据数理统计原理中的大数定律，足够大的样本的统计分布一般服从正态分布。而当样本数据服从正态分布时，该类数据的异常值检验、剔除可以采用拉伊达准则。

拉伊达准则是以最大误差范围 3σ 为依据进行判别，设有一组试验数据 $x_i(i=1, 2, \dots, n)$ ，其标准偏差为：

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\Delta x_i)^2} \quad (1)$$

式中： \bar{x} 为样本平均值； $\Delta x_i = x_i - \bar{x}$ 为偏差。如果某测量值 $x_i(i=1, 2, \dots, n)$ 的偏差 $|\Delta x_i| > 3\sigma$ 时，则认为 x_i 是含有过失误差的异常值。

拉伊达准则简单、方便，无需查数据表，适用于指控装备试验次数较多或要求不高的场合，当试验次数较少时不能应用。

利用拉伊达准则判断样本 $x_i(i=1, 2, \dots, n)$ 中是否含有异常值的步骤如下：

- 1) 计算样本均值 \bar{x} 与标准偏差 σ 。
- 2) 对偏差最大的数据采用拉伊达准则进行判断。如果该数据不属于过失误差，判断结束；如果该数据含有过失误差，则剔除该数据，并对剩下的 $n-1$ 个数据重新进行判断。
- 3) 计算剩下的 $n-1$ 个数据的样本均值 \bar{x}' 与标准偏差 σ' 。
- 4) 对剩下的 $n-1$ 个测量数据中偏差最大的数据再按拉伊达准则进行判断。这样一直进行下去，直到找不到含有过失误差的试验数据为止。

3.2 大样本数据统计分析

异常值剔除后，通常采用点估计或区间估计的

方法进行统计分析。点估计是在抽样推断中不考虑抽样误差，直接以抽样指标代替全体指标。区间估计是抽样推断中根据抽样指标和抽样误差估计全体指标的可能范围。对于大样本条件下的试验数据通常采用经典的数理统计方法进行处理，不再赘述。

4 小样本试验数据的异常值检验与统计

4.1 小样本数据异常值检验

在指控装备试验中，部分情况下，单次试验只能获得一组或少量数据。例如：比率、概率类数据，数量类数据，能力、质量、满意度类数据，最值类数据和直接观测类数据；因此，上述试验数据可能为小样本数据。对于小样本数据，其异常值更容易引起大的误差。一般采用如下特殊方法进行异常值检验。

1) 格拉布斯准则。

对于试验数据 $x_i(i=1, 2, \dots, n)$ ，定义统计量：

$$g_i = |\Delta x_i| / \sigma \quad (2)$$

式中： $\Delta x_i = x_i - \bar{x}$ 为偏差； σ 为标准偏差；选定显著性水平 α (常取值为 0.05、0.025、0.01)，如某试验数据 x_i 对应的 g_i 满足 $g_i > g_\alpha(n)$ ，则认为在显著性水平 α 时，该试验数据含有异常值，应予以剔除； $g_\alpha(n)$ 为格拉布斯临界值，可以从标准正态分布概率表中查到。

采用格拉布斯准则判断数据 $x_i(i=1, 2, \dots, n)$ 中是否有异常值的步骤如下：

- ① 首先将试验数据按从小到大的顺序排列，得到 $x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq x_n$ 。
- ② 计算样本均值 \bar{x} 与标准偏差 σ 。
- ③ 根据试验次数 n 和选定的显著性水平 α ，查出格拉布斯临界值 $g_\alpha(n)$ 。
- ④ 对与均值 \bar{x} 偏差最大的数据 $x_i(x_1$ 或者 $x_n)$ 进行判断。如果 x_i 对应的 $g_{(i)} > g_\alpha(n)$ ，则该数据不含过失误差，判断结束；如果 $g_i > g_\alpha(n)$ ，则该数据含有过失误差，应剔除该数据，并对剩下的 $n-1$ 个试验数据重新进行判断。

⑤将剩下的 $n-1$ 个试验数据按从小到大的顺序排列, 得到 $x'_1 \leq x'_2 \leq \dots \leq x'_{n-1}$, 计算剩下的 $n-1$ 个试验数据的样本均值 \bar{x}' 与标准偏差 σ' 。

⑥根据试验次数 $n-1$ 和选定的显著性水平 α , 查出格拉布斯临界值 $g_\alpha(n-1)$ 。

⑦对剩下的 $n-1$ 个试验数据中与 \bar{x}' 偏差最大的数据 x'_i (x'_1 或者 x'_{n-1}) 进行判断。如果 x'_i 对应的 $g'_i \leq g_\alpha(n-1)$, 则该数据不含过失误差, 判断结束; 如果 $g'_i > g_\alpha(n-1)$, 则该数据含有过失误差, 应剔除该数据。对剩下的 $n-2$ 个数据继续进行判断, 这样一直进行下去, 直到找不到含有过失误差的试验数据为止。

2) 狄克逊准则。

采用狄克逊准则可以避免计算样本均值和标准偏差。该准则采用了极差比的方法。为提高判断效率, 不同的试验次数应用不同的极差比计算方法。将试验数据按从小到大的顺序排列, 得到 $x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq x_n$ 。

如果上述试验数据中含有过失误差, 首先值得怀疑的是 x_1 和 x_n 。狄克逊准则定义了一个与 x_1 或 x_n 有关的极差比统计量 d_0 , 其计算公式如表 2 所示。

表 2 极差比统计量计算公式

n 的取值范围	极差比统计量 d_0 的计算公式	
	x_1 为异常值时	x_n 为异常值时
3~7	$(x_2-x_1)/(x_n-x_1)$	$(x_n-x_{n-1})/(x_n-x_1)$
8~10	$(x_2-x_1)/(x_{n-1}-x_1)$	$(x_n-x_{n-1})/(x_n-x_2)$
11~13	$(x_3-x_1)/(x_{n-1}-x_1)$	$(x_n-x_{n-2})/(x_n-x_2)$
14~25	$(x_3-x_1)/(x_{n-2}-x_1)$	$(x_n-x_{n-2})/(x_n-x_3)$

如果 $d_0 > g_\alpha(n)$, 则认为在显著性水平 α 下, x_1 (或 x_n) 中含有过失误差, 应予以剔除。 $g_\alpha(n)$ 为狄克逊检验临界值, 与试验数据的数量 n 、显著性水平 α 及 d_0 的计算公式有关, 可通过查阅狄克逊检验临界值表得到。

采用狄克逊准则判断试验数据 $x_i (i=1, 2, \dots, n)$ 中是否有异常值的步骤如下:

①首先将试验数据按从小到大的顺序排列, 得到 $x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq x_n$ 。

②对与均值 \bar{x} 偏差最大的数据 x_i (x_1 或者 x_n) 进行判断。根据试验次数 n 和选定的显著性水平 α 以及偏差最大的数据是 x_1 还是 x_n , 查出狄克逊检验临界值 $g_\alpha(n)$ 和 d_0 的计算公式。

③如果 x_i 对应的 $d_0 \leq g_\alpha(n)$, 则该数据不含过失误差, 判断结束; 如果 $d_0 > g_\alpha(n)$, 则该数据含有过失误差, 应剔除该数据。对剩下的 $n-1$ 个试验数据

重新按从小到大的顺序排列, 计算新的均值, 再重新进行判断。这样一直进行下去, 直到找不到含有过失误差的试验数据为止。

4.2 基于 Bootstrap 方法的小样本数据统计分析

针对小样本试验数据, 由于样本量不足, 利用传统统计方法获得的分析结果误差较大。需要采用合适的方法扩充试验数据样本容量, 然后再进行统计分析提取数据特征。Bootstrap 方法用原始试验数据样本自身的数据抽样得出新的样本及统计量, 一般称为“自助法”。它是一种有放回的抽样方法, 是非参数统计中一种重要的估计统计量方差的统计方法, 可用于指控装备小样本试验数据的统计分析。

设 $S_0 = \{x_1, x_2, \dots, x_n\}$ 是统计量 X 的 n 个原始试验数据。假设需要估计 $\theta = S(X)$ 的分布。Bootstrap 方法流程如图 2 所示。

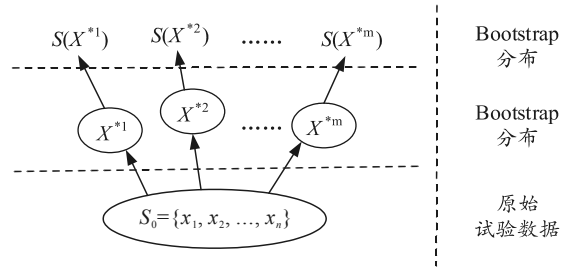


图 2 Bootstrap 方法流程

1) 从 S_0 中有放回地随机抽取 n 个样本, 称为一次随机抽取试验。得到 Bootstrap 样本 $X^* = \{x_1^*, x_2^*, \dots, x_n^*\}$, 称为一个自助样本。

2) 用 $\hat{\theta}^* = S(X^*)$ 的自助分布近似估计 $\theta = S(X)$ 的样本分布。

3) 对 S_0 进行 m 次随机抽取试验, 生成 m 个相互独立的自助样本 $X^{*1}, X^{*2}, \dots, X^{*m}$ 。

4) 对每个 X^{*i} 求解 $\hat{\theta}$ 的估计值:

$$\hat{\theta}^*(i) = S(X^{*i}), \quad i=1, 2, \dots, m. \quad (3)$$

5) 计算 Bootstrap 估计的标准误差, 即 $\hat{\theta}^*(i)$ 的标准差:

$$\hat{e}_m = \sqrt{\frac{1}{m-1} \sum_{i=1}^m [\hat{\theta}^*(i) - \hat{\theta}^*(\cdot)]^2}. \quad (4)$$

其中 $\hat{\theta}^*(\cdot) = \frac{1}{m} \sum_{i=1}^m \hat{\theta}^*(i)$ 。

5 结束语

笔者针对指控装备效能评估的军事需求及试验数据分析的现状, 分析了指控装备试验数据的类型

构成，提出需要区分试验数据样本量大小，分别进行异常值检验、剔除，再进行统计分析。特别是当前指控装备更新换代快、试验数据不足等导致的小样本试验数据分析需求非常强烈。笔者建议采用 Bootstrap 方法自助扩充试验数据样本容量，再对统计量方差进行估计。同时，试验数据异常值检验和统计分析还应注意以下几点：

1) 检验多个异常值数据时，异常值数据应逐一检验，主要考虑不同数据的异常程度是不一致的，应按照偏差从大到小的顺序来检验。如按照检验准则该数据不应被剔除，则所有偏差更小的数据都不应被剔除，也无需再检验。剔除一个数据后，再检验下一个数据时，应注意试验数据的总数量发生了变化。

2) 用不同的方法检验同一组试验数据，在相同的显著性水平上，可能会产生不同的结论。对于要求较为精确的试验，可以选用 2~3 种方法对试验数据进行判断。

3) 在异常值剔除过程中，应当对每个剔除的数据进行误差分析，开展粗大误差判别，以决定数据的取舍。对待粗大误差，除通过从测量结果中及时发现和鉴别外，还需要提高测量的技术水平和责任心，保证测量条件的稳定，提高测量的精度。

4) 在异常值检验、剔除后，进行数据统计分析时，采用 Bootstrap 自助抽样法对小样本数据的统计分析非常有效，该方法在大样本数据下也可取得较好的效果；但对于极小样本的处理将会产生一定偏差。经验表明，初始小样本的数量下限为 5，一般最好大于 15，自助样本数量一般应大于 1 000^[11-12]。

(上接第 40 页)

- [6] 高薇, 宋凌峰, 王祝堂. 中国客运轨道车体铝型材的深加工[J]. 轻合金加工技术, 2016, 44(1): 1-11.
- [7] 蓝剑红. CRH3 型动车组车体刚度有限元分析[D]. 成都: 西南交通大学, 2012.
- [8] 梁建光. 飞机蒙皮切边柔性夹具夹持方案的优化研究[D]. 上海: 上海交通大学, 2013.

参考文献:

- [1] 王金良, 郭齐胜, 李玉山, 等. 陆军开展装备作战试验的总体思考[J]. 装甲兵工程学院学报, 2016, 30(3): 1-6.
- [2] 夏杰, 田傲宇, 王栋. 装备作战试验数据分析与处理方法[J]. 中国设备工程, 2020(2): 160-161.
- [3] 李志强, 汪新, 陈宇奇, 等. 基于小样本失效数据的复杂装备可靠性评估研究进展与挑战[J]. 航空兵器, 2021, 28(3): 83-90.
- [4] 方志耕, 陈顶, 刘思峰. 贫信息背景的复杂装备可靠性预测现状与展望[J]. 指挥信息系统与技术, 2018, 9(5): 1-8.
- [5] 王璇, 狄鹏. 基于虚实试验数据融合的装备测试性评估方法[J]. 舰船电子工程, 2021, 41(6): 131-134.
- [6] 李文丽, 原大宁. 极小子样下参数随机分布的电主轴可靠性分析[J]. 机械科学与技术, 2016, 35(5): 690-695.
- [7] 叶伟, 李亚平, 陈顶, 等. 基于小样本的复杂装备“虚拟总体”生成与检验模型[J]. 中国管理学, 2015, 23(s1): 240-244.
- [8] 张春丽, 杨利胸, 李庆男. 门槛存在性 Bootstrap 检验有限样本性质的蒙特卡洛模拟分析[J]. 统计与决策, 2019(4): 37-40.
- [9] 刘金娣, 李莉莉, 高静, 等. 异常值检验方法的比较分析[J]. 青岛大学学报, 2017, 30(2): 106-109.
- [10] 李亚楠, 田雪颖, 王志梅. 基于大数据的航天装备试验鉴定数据管理及分析应用研究[J]. 航天工业管理, 2018(10): 35-40.
- [11] 王丙参. Bootstrap 计失效情形的探讨[J]. 统计与决策, 2018, 20(16): 71-74.
- [12] 刘思雨, 张敏强. Bootstrap 平均数假设检验样本容量探讨[J]. 统计与决策, 2016(14): 26-28.
- [9] FUH J, DEVORR E, KAPOORS G. A Mechanistic Model for the Predication of the Force System in Face Milling Operation[J]. Journal of Engineering for Industry, 2018, 106(1): 81-88.
- [10] SUBRAMANI G, SUVADA R, KAPOOR S G. A Model for the Predication of Force System for Cylinder Boring Process[J]. Transactions of NAMRI/SME, 2017, 2(1): 439-446.