

齿轮测量中异常值的判定及剔除系统设计

时海涛, 苏智剑

(郑州大学 机械工程学院, 河南 郑州 450001)

摘要:针对齿轮测量数据中存在异常值的问题,分别对异常值的4种常用准则进行了对比分析;设计了异常值剔除系统,结合齿轮测量系统所测得的齿轮数据,选用相应的判定准则,对齿轮测量数据进行快速高效的异常值剔除处理。通过实验分析,给出了齿轮测量数据预处理需要遵循的原则。研究表明,当对齿轮测量数据进行粗大误差剔除时,需要结合实际情况对4种准则综合运用,以达到所得数据的科学性和合理性,对齿轮误差的后续处理做出科学的预处理。

关键词:粗大误差;异常值;剔除;判定准则

中图分类号:TH132.41 **文献标志码:**B **文章编号:**1671-5276(2020)03-0053-04

Design of Judgment and Elimination System for Outliers in Gear Measurement

SHI Haitao, SU Zhijian

(School of Mechanical Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: For the outliers existing in gear measurement data, the four criteria commonly used for the outliers are compared and analyzed. The outlier elimination system is designed with the gear data measured by the gear measurement system. The corresponding determination criteria are selected to perform fast and efficient outlier elimination processing on the gear measurement data. Through the experimental analysis, the principles that need to be followed in the preprocessing of gear measurement data are given. Because the gear measurement is a necessary work, the measurement data is to be precise, and the experimental results obtained by the four criteria are carefully studied and analyzed. The results of the analysis show that when the gear measurement data is grossly eliminated, it is necessary to comprehensively use the four criteria with the actual situation to achieve the science and rationality of the obtained data, and to make a scientific pretreatment for the subsequent analysis and processing of the gear error.

Keywords: gross error; outlier; elimination; determination criterion

0 引言

齿轮误差的测量是一项要求十分严格的工作,测量数据的处理分析是齿轮误差测量实验的关键一环,它直接关系到对齿轮制造精度以及齿轮测量装置的稳定性和可靠性的评价。在齿轮现场测量的过程中,由于测量环境的不稳定、操作人员操作不当等因素都可能会造成测量时测量数据的突变,进而产生异常值,即粗大误差^[1]。测量数据的预处理是为了剔除由上述因素产生的粗大误差,从而保证测量数据的科学性和准确性^[2]。

本文在对4种粗大误差判定准则分析研究的基础上设计了粗大误差判定系统^[3],对齿轮测量数据进行异常值处理以保证测量数据的可靠性,以便对齿轮的误差作出科学分析,为齿轮在装配使用时预先做出装配方案提供科学的依据。

1 测量数据异常值的判定方法

测量数据中的粗大误差是一组测量数据出现的明显

脱离其他正常值的那些异常值,这些数值相差比较大,而且离散出现^[4]。常用的判定准则有狄克逊准则、拉伊达准则、格拉布斯准则以及肖维勒准则。在齿轮误差测量的过程中,粗大误差需要慎重处理,要结合其判定准则作出科学的分析,综合运用多种判定准则处理异常值,以避免粗大误差的遗漏和测量真值的误判。

1.1 狄克逊判别法

狄克逊准则是运用极差比法定义了4个统计量,根据测量数据量的不同分别采用不同的统计量。在计算统计量之前,需要先对测量数据进行从小到大的排序。有下面4个统计量:

$$\begin{aligned} r_{10} &= \frac{x_n - x_{n-1}}{x_n - x_1}; r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}; \\ r_{21} &= \frac{x_n - x_{n-2}}{x_n - x_2}; r_{22} = \frac{x_n - x_{n-1}}{x_n - x_3} \end{aligned} \quad (1)$$

当 $3 \leq n \leq 7$ 时,用统计量 r_{10} 进行计算判别;当 $8 \leq n \leq 10$ 时,用统计量 r_{11} 进行计算判别;当 $11 \leq n \leq 13$ 时,用统计量 r_{21} 进行计算判别;当 $14 \leq n \leq 15$ 时用统计量 r_{22} 进行计算判别。通过查询临界系数 $D(n, \alpha)$ 表,如果

$$r_{ij} > D(n, \alpha) \quad (2)$$

那么判定 x_n 或 x_1 为粗大误差值, 否则为正常值。临界系数 $D(n, \alpha)$ 存于 excel 表中; 当用狄克逊法判定时, 判定系统直接从 excel 表中调用 $D(n, \alpha)$ 进行计算判别。其中 n 表示数据量, x_n 为第 n 个数据, α 为置信度, 一般取置信度 α 为 0.01。部分 $D(n, \alpha)$ 值如表 1 所示。

表 1 部分狄克逊准则临界系数 $D(n, \alpha)$

数据量 n	$D(n, \alpha)$	
	$\alpha = 0.05$	$\alpha = 0.01$
3	0.970	0.994
4	0.829	0.926
5	0.710	0.821
6	0.628	0.740
7	0.569	0.680
8	0.608	0.717
9	0.564	0.672
10	0.530	0.635
11	0.619	0.709
12	0.583	0.660
13	0.557	0.638
14	0.587	0.669
15	0.565	0.646
16	0.547	0.629
17	0.527	0.614
18	0.513	0.602

1.2 拉伊达判别法

拉伊达准则适用于测量数据量较大的情况, 又称为 3σ 准则; 当数据量 $n \leq 10$ 时, 此方法无效^[5]。设 X_n 为一组实验数据, X_i 为可疑值, 若要判断出这组数据中的粗大误差值, 则需要先求出这组数据的算术平均值 $\bar{X}$ 及标准偏差 σ , 若

$$|X_i - \bar{X}| > 3\sigma \quad (3)$$

则判定 X_i 为粗大误差, 否则为正常值。

当采用拉伊达准则判定粗大误差值时, 每剔除一个粗大误差值就需要重新计算剩余数据的算术平均值和标准偏差, 依据判定准则再次判定剩余数据中的粗大误差, 如此循环往复直至剔除全部误差值。

1.3 肖维勒判别法

设 X_n 为一组实验数据, $\bar{X}$ 为此组测量数据的算术平均值, σ 为其标准偏差, k 为肖维勒系数, X_i 为可疑值; 若

$$|X_i - \bar{X}| > \sigma > k \quad (4)$$

则判定 X_i 为粗大误差值, 否则为正常值。剔除粗大误差值后, 重新计算剩余数据的平均值和标准偏差, 依次用判别法判别粗大误差, 如此循环直至剔除全部误差。其中肖维勒系数存于 excel 表中。

1.4 格拉布斯判别法

格拉布斯判别法通常适用于小样本容量的情况, 在用格拉布斯判别法判定粗大误差值之前, 要先将测量数据按从小到大的顺序排列^[6], 然后构造统计量 G_1 和 G_n , 并与相应置信度 α 的统计临界系数 $G(\alpha, n)$ 的值进行比较, 最终判断出粗大误差值; 当判断出粗大误差值后, 把剩余数据按以上方法重复进行判别, 直至剔除所有粗大误差值, 具体方法如下:

$$G_1 = (X - X_1) / \sigma, G_n = (X_n - X) / \sigma \quad (5)$$

当 $G_1 \geq G_n$ 且 $G_1 > G$ 时, 则 X_1 为粗大误差值; 当 $G_n \geq G_1$ 且 $G_n > G_0$ 时, 则 X_n 为粗大误差值; 当 $G_1 < G_0$ 且 $G_n < G_0$ 时, 则不存在粗大误差值。其中 $\bar{X}$ 为测量数据的算术平均值, σ 为标准偏差, G_0 为相应的临界系数值, 通常取置信度 α 为 0.05。部分 $G(\alpha, n)$ 的值如表 2 所示。

表 2 部分格拉布斯统计临界系数 $G(\alpha, n)$

数据量 n	$G(\alpha, n)$			
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.005$
3	1.153	1.155	1.155	1.155
4	1.463	1.492	1.481	1.496
5	1.672	1.749	1.715	1.764
6	1.822	1.944	1.887	1.973
7	1.938	2.097	2.020	2.139
8	2.032	2.220	2.126	2.274
9	2.110	2.323	2.215	2.387
10	2.176	2.410	2.290	2.482
11	2.234	2.485	2.355	2.564
12	2.285	2.550	2.412	2.636
13	2.331	2.607	2.462	2.699
14	2.371	2.659	2.507	2.755
15	2.409	2.705	2.549	2.806
16	2.443	2.747	2.585	2.852
17	2.475	2.785	2.620	2.894
18	2.501	2.821	2.651	2.932

2 异常值判剔除系统设计

本文在对 4 种粗大误差判定准则分析的基础上, 基于 VB6.0 开发设计了齿轮测量异常值剔除系统; 该系统能够在齿轮测量时, 对由于操作人员操作不当或其他外界突发等状况而引起的测量异常值进行及时发现和剔除^[7], 以便对齿轮的某项数据进行重新精确的测量或补偿^[8]。

2.1 系统结构

齿轮测量粗大误差剔除系统主要有数据读入、误差剔除、数据查询、保存、刷新、帮助几个部分组成, 系统界面如图 1 所示。

2.2 主要模块及原理

1) 数据输入、输出模块



图1 齿轮测量异常值剔除系统界面

原始测量数据的读入以及处理后数据的输出、保存都采用 I/O 接口模块来实现。在实际测量过程中,齿轮测量数据量是比较大的,如果用人工输入的方式则会浪费大量的时间,严重降低测量工作的效率,所以本文采用文本格式(*.txt)的方式进行数据的读入及输出,大大地提高了数据处理的效率。

2) 误差剔除模块

误差剔除模块由拉伊达、格拉布斯、狄克逊、肖维勒 4 个判别子模块组成,在进行误差剔除时点击相应菜单下的判别法进行数据判别处理。

3) 数据库访问模块

本系统除了拉伊达判别法外,其他 3 种判别法在进行粗大误差判定时,依据测量数据量的不同都要重复调用相应的判据系数,所以本文运用数据库访问技术,即 ADO 技术,对判据系数 excel 数据库表进行访问^[9],从而大大提高了数据处理的效率。以下为部分 excel 数据库访问程序代码:

```
Public Sub shujuku_Click()  
Dim cn As New ADODB.Connection  
Dim rs As New ADODB.Recordset  
cn.ConnectionString = "Provider=Microsoft.Jet  
.OLEDB.4.0;PersistSecurityInfo=false;  
DataSource = "&App.Path&"\form\shujuk  
u.xls;Extended  
Properties = 'Excel8.0;HDR=Yes'  
'HDR=Yes  
cn.Open  
rs.Open "select * from[ sheet1 $ ]", cn,  
adOpenKeyset, adLockOptimistic  
If Not rs.EOF Then rs.MoveFirst  
rs.Move (n)  
For i = 0 To n  
rs.MovePrevious  
o = rs.Fields.Item(0).Value  
If Option1.Value = True Then  
r0 = rs.Fields.Item(2).Value  
ElseIf Option2.Value = True Then  
r0 = rs.Fields.Item(1).Value  
End If
```

3 齿轮测量数据的异常值剔除

依据上述设计的异常值剔除系统,对某齿轮的齿距误

差测量数据进行粗大误差的剔除。该齿轮的设计齿数 $z=48$,模数 $m=4$,分度圆压力角 $\alpha=20^\circ$ 。齿轮齿距偏差的测量属于精密测量的范畴,所以选用多种判别法进行综合判断,当判别时运行结果相同时,则认为某测量数据应该剔除,否则需慎重处理。

以该齿轮所测量的 48 个齿距偏差数据为实验数据,取不同数量的测量数据用 4 种判别法依次进行异常值的判定和剔除。表 3 为该齿轮齿距偏差测量数据。

表 3 齿轮 48 个齿距偏差测量值 单位: μm

编号	齿距偏差	编号	齿距偏差
1	13.5	25	-7.1
2	17.3	26	26.6
3	24.7	27	19.5
4	-16.2	28	29.1
5	-6.7	29	17.7
6	-22.1	30	-9.8
7	22.6	31	14.4
8	11.5	32	-26.3
9	25.2	33	10.4
10	-1.6	34	97.6
11	1.1	35	26.1
12	-9.6	36	31.5
13	15.9	37	-13.8
14	6.4	38	16.4
15	19.7	39	8.1
16	-4.3	40	-3.6
17	-84.3	41	11.5
18	25.6	42	112.5
19	21.2	43	19.6
20	-0.7	44	7.5
21	22.3	45	31.4
22	-16.4	46	28.8
23	22.1	47	-5.3
24	27.6	48	-11.7

3.1 肖维勒判别法实验

根据第 2 节对 4 种粗大误差判别准则的分析,选用肖维勒准则对 48 个齿轮偏差数据进行处理,系统运行结果如图 2 所示。

由系统运行结果可以看出,在 48 个齿距偏差数据中,有 3 个齿距偏差数据-84.3、112.5、97.6 是异常的,即为粗大误差值。

3.2 格拉布斯判别法实验

齿距偏差数据为 48 个时,选用格拉布斯准则且置信度 $\alpha=0.05$ 时进行异常值的判定与剔除,没有判别出粗大误差。取齿距偏差中的 23 个齿距偏差数据进行判定,运行结果如图 3 所示。

当用格拉布斯判别法进行判别时,从运行结果可以看出,在对混入 3 个粗大误差的相同 48 个齿距偏差数据进行处理时,没有判断出粗大误差;当取 21 个齿距偏差,其中存在 1 个异常值的数据进行判别时,检测出粗大误差值为-84.3。

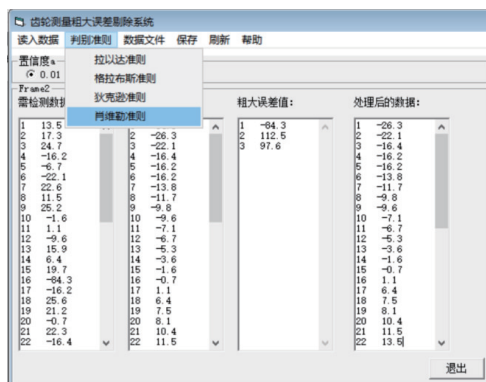


图2 肖维勒法判别运行结果

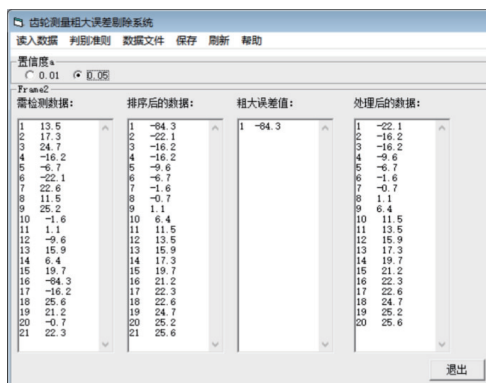


图3 格拉布斯法判别运行结果

3.3 拉伊达判别法实验

当使用拉伊达法对含有3个异常值的48个齿距偏差数据进行判定时,检测出3个异常值,分别为-84.3、112.5、97.6。当对其中含有1个异常值的11个数据进行判定时没有判定出异常值,运行结果如图4所示。

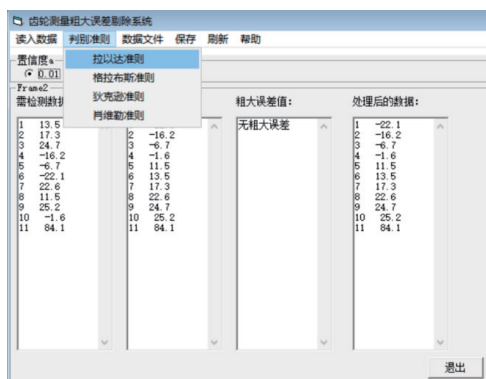


图4 拉伊达法判别运行结果

3.4 狄克逊判别法实验

当使用狄克逊法对含有3个异常值的48个齿距偏差进行检测,判别出3个粗大误差值,其运行结果同拉伊达判别法;当对其中10个齿距偏差数据进行判别时,剔除1个含有较大误差的正常齿距偏差测量值。

3.5 测量结果分析

通过粗大误差判定系统的运行结果分析,可以得到4种判别法的特点及适用性。拉伊达法和肖维勒判别法在数据量比较大时,可以快速判别出粗大误差值;狄克逊判别法可以一次剔除多个异常值;当1组测量数据中含有1个异常值时,适合用格拉布斯法进行异常值的判别与剔除。

在齿轮误差测量的过程中,要求具有较高的测量效率以及较高测量精度,所以在进行齿轮测量数据进行异常值判别及剔除时,可以用多种方法对同一组齿轮误差数据进行异常值的判定,如可以分别使用拉伊达准则和狄克逊准则对误差数据进行异常值的判定剔除,当判定结果相同时,可以认定为异常值,否则要对检测数据慎重处理;数据量较少时,使用前述方法判定后,可以用格拉布斯判别法对数据再次进行精确的判定。

4 结语

齿轮误差的分析评价是建立在齿轮误差实际测量数据基础之上的。在对齿轮误差数据进行异常值判定时,要综合运用多种判别准则进行异常值的判定及剔除,以保证齿轮误差数据的稳定性和精确性。通过对4种判别法的分析,以及用设计的粗大误差剔除系统对齿轮齿距偏差的数据预处理实验,验证了4种判别法在齿轮齿距偏差数据预处理中的科学性以及适用性。对齿轮测量数据预处理方法的研究,进一步完善了齿轮测量数据误差的处理方法;在齿轮测量过程中,可以有效地避免因为人为因素或者环境因素而掺入异常测量值,保证后续齿轮误差分析评价工作中所用数据的有效性和科学性。

参考文献:

- [1] 于晓峰. 装备实验数据处理[M]. 北京:国防工业出版社, 2010:1-15.
- [2] 费业泰. 误差理论与数据处理[M]. 北京:机械工业出版社, 2005.
- [3] 叶川,伍川辉,张嘉怡. 计量测试中异常数据剔除方法比较[J]. 计量与测试技术,2007,34(7):26-27.
- [4] 王鹏,王健安,郭畅,等. 基于云计算及数据挖掘技术的海量数据处理研究[J]. 长春理工大学学报(自然科学版),2013,36(6):157-160.
- [5] 韩旭,王建宇,祖先锋. 基于时间序列模型的系统最大值指标评定方法[J]. 系统工程与电子技术,2012,34(4):839-845.
- [6] 刘建,刘文金. 应用格罗布斯准则判定测量结果中的粗大误差[J]. 设计与研究,2006(2):20-21.
- [7] 李梦奇,白晓军. 工程技术领域等精度数据异常值判定系统[J]. 计量技术,2006(3):57-59.
- [8] 邱林,宋爱平,沈雨涵,等. 基于激光位移传感器的锥齿轮齿距偏差及齿圈跳动测量[J]. 机械传动,2017(11):174-177.
- [9] 林卓然. VB 语言程序设计[M]. 北京:电子工业出版社, 2003.

收稿日期:2019-02-19