



中华人民共和国国家标准

GB/T 6380—2008
代替 GB/T 6380—1986

数据的统计处理和解释 I 型极值分布样本离群值的判断和处理

Statistical interpretation of data—Detection and treatment of
outliers in the sample from type I extreme value distribution

2008-07-28 发布

2009-01-01 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

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语、定义和符号	1
3.1 术语和定义	1
3.2 符号和缩略语	2
4 离群值判断	2
4.1 离群值的来源与判定	2
4.2 检出离群值个数的上限	2
4.3 单个离群值情形	2
4.4 多个离群值情形	2
5 离群值处理	3
5.1 处理方式	3
5.2 处理规则	3
5.3 备案	3
6 单个离群值的判断规则	3
6.1 检验方法的选择	3
6.2 狄克逊(Dixon)检验法	3
6.3 欧文(Irwin)检验法	4
7 多个离群值的判断规则	5
7.1 检验步骤	5
7.2 多个离群值检验示例	5
附录 A (规范性附录) 临界值表	6
参考文献	8

前 言

本标准代替 GB/T 6380—1986《数据的统计处理和解释 I 型极值分布样本异常值的判断和处理》。本标准与 GB/T 6380—1986 相比较,技术内容的变化主要包括:

- 按 GB/T 1.1—2000《标准化工作导则 第 1 部分:标准的结构和编写规则》的要求对标准格式进行了修改;
- 增加了术语、定义和符号;
- 将标准名称中“ I 型极值分布样本异常值的判断和处理”改为“ I 型极值分布样本离群值的判断和处理”;
- 将术语“检出异常值”和“高度异常值”分别改为“歧离值”和“统计离群值”,并进一步明确了二者的含义及相互差异;
- 增加了检出水平和剔除水平的定义;
- 检出水平由原标准中“检出水平一般取为 1%,5%或 10%”改为“除非根据本标准达成协议的各方另有约定,检出水平应为 0.05”;
- 明确规定剔除水平的值,除非根据本标准达成协议的各方另有约定,剔除水平应为 0.01;
- 增加了各种情形“统计离群值”的检验步骤;
- 增加了“多个离群值的判断规则”;
- 将“没有异常值”和“没有高度异常的异常值”分别改为“未发现离群值”和“未发现统计离群值”。

本标准的附录 A 是规范性附录。

本标准由全国统计方法应用标准化技术委员会提出并归口。

本标准起草单位:宁波工程学院、中国标准化研究院、北京大学、天津大学、海南省产品质量监督检验所。

本标准主要起草人:荆广珠、丁文兴、于振凡、蔡军伟、孙山泽、马逢时、黄艳等。

本标准所代替标准的历次版本发布情况为:

- GB/T 6380—1986。

引 言

从事科学研究、工农业制造以及管理工作都离不开数据,而对这些数据的整理、分析和解释都离不开统计方法。统计学是研究数字资料的整理、分析和正确解释的一门学科。人们各自从不同的来源取得各种数字资料,这些数字资料通常都是杂乱无章的,必须经过整理和简缩才能利用,使用完善的统计方法就可使数据整理、排列的有条有理,用图形或少量的几个重要参数,就可将大量数据的特征表达出来,这样既可避免不正确的解释,又可将获得满意数据的成本降到最低限度,提高了经济效益。

国家标准《数据的统计处理和解释》包含以下各项:

- 统计容忍区间的确定(GB/T 3359)
- 均值的估计和置信区间(GB/T 3360)
- 在成对观测值情形下两个均值的比较(GB/T 3361)
- 二项分布参数的估计与检验(GB/T 4088)
- 泊松分布参数的估计和检验(GB/T 4089)
- 正态性检验(GB/T 4882)
- 正态样本离群值的判断和处理(GB/T 4883)
- 正态分布均值和方差的估计与检验(GB/T 4889)
- 正态分布均值和方差检验的功效(GB/T 4890)
- I型极值分布样本离群值的判断和处理(GB/T 6380)
- 伽玛分布(皮尔逊Ⅲ型分布)的参数估计(GB/T 8055)
- 指数分布样本离群值的判断和处理(GB/T 8056)

本标准尚无相应的国际标准。

数据的统计处理和解释

I 型极值分布样本离群值的判断和处理

1 范围

本标准规定了判断和处理 I 型极值分布的样本中上侧离群值和 I 型极小值分布的样本中下侧离群值的一般原则和实施方法。

本标准适用于来自 I 型极值分布或 I 型极小值分布总体的样本。

本标准适用于样本量为 5 至 50 的情形。

注：由于经过变换 $Z = -X$ 后，I 型极小值分布的随机变量将服从 I 型极值分布，所以只对 I 型极值分布给出检出上侧离群值的方法。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。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标准，然而，鼓励根据本标准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标准。

ISO 3534-1:2006 统计学词汇及符号 第 1 部分：一般统计术语与用于概率的术语

ISO 3534-2:2006 统计学词汇及符号 第 2 部分：应用统计

3 术语、定义和符号

ISO 3534-1:2006、ISO 3534-2:2006 确立的术语、定义和符号以及下列术语、定义和符号适用于本标准。

3.1 术语和定义

3.1.1

I 型极值分布 type I extreme value distribution (Gumbel distribution)

具有下述分布函数的连续型分布。

$$F(x) = \exp(-e^{-(x-a)/b})$$

其中： $b > 0$ ， $-\infty < x < +\infty$

3.1.2

I 型极小值分布 type I minimum value distribution

具有下述分布函数的连续型分布。

$$F(x) = 1 - \exp(-e^{-(x-a)/b})$$

其中： $b > 0$ ， $-\infty < x < +\infty$

3.1.3

离群值 outlier

样本中的一个或几个观测值，它们离开其他观测值较远，暗示它们可能来自不同的总体。

注：离群值按显著性的程度分为歧离值和统计离群值。

3.1.4

统计离群值 statistical outlier

在剔除水平(3.1.7)下统计检验为显著的离群值。