



中华人民共和国国家标准

GB/T 6380—2019
代替 GB/T 6380—2008

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

2019-12-10 发布

2020-07-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言	I
引言	II
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	3
4 离群值判断	3
4.1 离群值的来源与判定	3
4.2 检出离群值个数的上限	3
4.3 单个离群值情形	3
4.4 多个离群值情形	4
5 离群值处理	4
5.1 处理方式	4
5.2 处理规则	4
5.3 备案	4
6 单个离群值的判断规则	4
6.1 检验方法的选择	4
6.2 狄克逊(Dixon)检验法	4
6.3 欧文(Irwin)检验法	6
7 多个离群值的判断规则	7
7.1 检验步骤	7
7.2 多个离群值检验示例	7
附录 A (规范性附录) 狄克逊检验的临界值表	8
附录 B (规范性附录) 欧文检验的临界值表	9
参考文献	10

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准代替 GB/T 6380—2008《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》，与 GB/T 6380—2008 相比，主要技术内容变化如下：

- 修改了术语 I 型极值分布(见 3.1.1,2008 年版的 3.1.1)；
- 修改了术语 I 型极小值分布(见 3.1.2,2008 年版的 3.1.2)；
- 修改了术语离群值(见 3.1.3,2008 年版的 3.1.3)；
- 增加了次序统计量的定义及说明(见 3.1.8)；
- 修改了离群值的处理规则(见 5.2,2008 年版的 5.2)；
- 增加了威布尔分布(Ⅲ型极小值分布)取对数后转化为 I 型极小值分布的说明及离群值计算示例(见 6.2.3)；
- 增加了参考文献 ISO 16269-4:2010(见参考文献)。

本标准由全国统计方法应用标准化技术委员会(SAC/TC 21)提出并归口。

本标准起草单位：天津大学、北京理工大学珠海学院、宁波工程学院。

本标准主要起草人：马逢时、许其州、史道济、荆广珠。

本标准所代替标准的历次版本发布情况为：

- GB/T 6380—1986、GB/T 6380—2008。

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,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、排列得有条有理,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。

根据收集到的样本观测值,可以大体上确认样本来自为某种分布。这时发现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。是否为实际的离群值,需要统计检验。在离群值的检验中特别要注意数据来自哪一种基本分布,假定来自 I 型极值分布与假定来自正态分布在检验时是非常不同的,对分布的错误假定会导致观测值被错误地归类为离群值。

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与检验(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);
- Γ 分布(皮尔逊 III 型分布)的参数估计(GB/T 8055);
- 指数样本离群值的判断和处理(GB/T 8056)。

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的上侧离群值和 I 型极小值分布的样本中的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,以及样本量为 5~50 的情形。

注 1: 由于经过变换 $Y = -X$ 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。

注 2: 由于 III 型极小值分布(威布尔分布)经过对数变换 $Z = \ln X$ 后将服从 I 型极小值分布,所以本标准实际上对 III 型极小值分布(威布尔分布)给出检出下侧离群值的方法。

注 3: I 型极值分布在水文、气象、地震、可靠性以及金融等很多领域有着广泛的应用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 3358.1 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

GB/T 3358.2 统计学词汇及符号 第 2 部分:应用统计

3 术语、定义和符号

3.1 术语和定义

GB/T 3358.1 和 GB/T 3358.2 界定的以及下列术语和定义适用于本文件。

3.1.1

I 型极值分布 type I extreme value distribution

具有下述分布函数

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

的连续型分布,其中 $b > 0$, $-\infty < a < \infty$, $-\infty < x < \infty$ 。

注 1: 当 $a=0, b=1$ 时, I 型极值分布的概率密度函数曲线图见图 1。

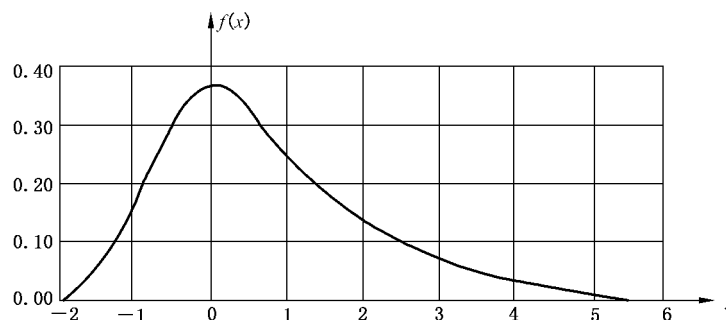


图 1 概率密度函数曲线图