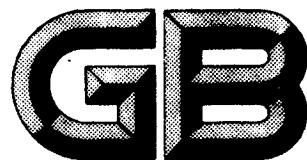


UDC 519.28
A 41



中华人民共和国国家标准

GB 8056—87

数据的统计处理和解释 指数样本异常值的判断和处理

Statistical interpretation of data
Detection and handling of outlying observations
in exponential sample

1987-07-08发布

1988-04-01实施

国家标准局 发布

数据的统计处理和解释
指数样本异常值的判断和处理

Statistical interpretation of data
Detection and handling of outlying observations
in exponential sample

1 目的和应用范围

1.1 本标准规定了在来自指数分布(单参数)的随机样本中判断和处理异常观测值的一般原则和实施方式。适用于来自指数总体或近似指数总体的样本,即除了个别或少数异常值外,其余大部分数据(主体数据)来自同一个指数总体或近似指数总体。指数分布的分布函数为:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases} \dots\dots\dots (1)$$

概率密度函数为:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases} \dots\dots\dots (2)$$

1.2 异常观测值(或称异常值)是指样本中的个别值,其数值明显偏离它(或它们)所在样本的其余观测值。

1.2.1 异常值可能仅仅是数据中固有的随机变异性的极端表现,若确是这样,就应把它和样本中其它观测值以同样的方式对待。

1.2.2 异常值也可能是偶然偏离所规定的试验条件和试验方法的后果。或是在计算或记录这个数值时出现的失误。这种异常值与其它观测值不属于同一个总体。

2 引用标准

- GB 3358—82 《统计学名词及符号》
- GB 4086.1~4086.6—83 《统计分布数值表》

3 符号及其意义

- $X_{(i)}$ —— 观测值中自小到大的第*i*个观测值。
- $T_{n(n)}$ —— 样本大小*n*<100时,检验最大的观测值 $X_{(n)}$ 是否为异常值时所用的统计量。
- $T_{n(1)}$ —— 样本大小*n*<100时,检验最小的观测值 $X_{(1)}$ 是否为异常值时所用的统计量。
- $E_{n(n)}$ —— 样本大小*n*>100时,检验最大的观测值 $X_{(n)}$ 是否为异常值时所用的统计量。
- $E_{n(1)}$ —— 样本大小*n*>100时,检验最小的观测值 $X_{(1)}$ 是否为异常值时所用的统计量。
- α —— 检验的显著性水平。
- $T_{n(n)}(1-\alpha)$ —— 显著性水平为 α 时,用统计量 $T_{n(n)}$ 作检验时的临界值。
- $T_{n(1)}(\alpha)$ —— 显著性水平为 α 时,用统计量 $T_{n(1)}$ 作检验时的临界值。
- $E_p(v_1, v_2)$ —— 自由度为 v_1 和 v_2 的*F*变量的*p*分位数。
- $E_{n,r(1)}$ —— 定数截尾样本中,判断 $X_{(1)}$ 是否为异常小时所用的统计量。