科学网《测量误差与小样本理论研讨会 - 2》
2022年7月28日

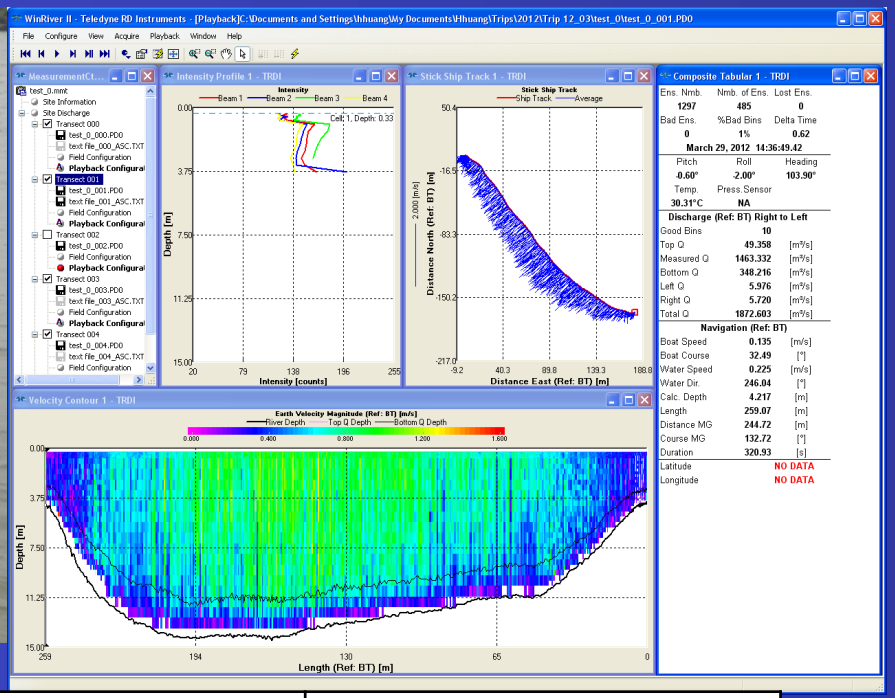
基于学生氏*t*-分布推断的谬误：两个悖论及其解决方法，及‘*t*-转换扭曲’

黄河宁
美国 Teledyne RD 仪器公司（已退休）

ADCP(声学多普勒流速剖面仪)河流流量测验涉及大量的小样本：样本量n = 2, 4, 6, …

例：泰国Choa Praya 河





测次	开始岸	开始时间	流量 (m^3/s)	测次流量与平均流量差 (%)
1	左岸	14:25:49	1873.2	-0.62
2	右岸	14:31:28	1872.6	-0.66
3	左岸	14:46:10	1901.0	0.85
4	右岸	14:51:26	1893.1	0.43
平均			1885.0	

为什么质疑学生氏*t*-分布?

根据《测量不确定度表示指南》，多次测量样本均值的扩展不确定度计算公式为

当总体标准差σ已知(B类不确定度评估)：

$$U_z = z_{95}\frac{\sigma}{\sqrt{n}}$$ -- 不确定度‘约定真值’(*z*-区间半宽)

当σ未知，用*s*替代(A类不确定度评估)：

$$U_t = t_{95}\frac{s}{\sqrt{n}}$$ -- 简称 *t* 不确定度(*t*-区间半宽)

式中：
z_{95} =对应于95%置信水平的 *z*-值
t_{95} =对应于95%置信水平的 *t*-值(自由度=*n*-1)
s =样本标准差

为什么质疑学生氏*t*-分布? (续)

水 利 学 报

2006 年 5 月 SHUILI XUEBAO 第 37 卷 第 5 期

文章编号:0559-9350(2006)05-0619-06

ADCP 流量测验随机误差分析 I :随机不确定度预测模型

黄河宁

(Teledyne RD Instruments, Inc.,圣地亚哥,美国)

将式(6)带入式(3),则当总体标准差未知时 M 次流量测验的随机不确定度可以由其平均值和标准差来计算

$$\varepsilon_Q = \frac{z_{\alpha/2}}{C\sqrt{M}} \frac{\sigma'_Q}{\overline{Q}} \quad M \geqslant 2 \tag{7}$$

式中:σ'_Q/Q 为相对标准差。

注意到当流量测验总体标准差未知时,没有采用学生氏 t 分布定义随机不确定度。这是因为流量测验随机误差服从正态分布的假定是合理的,如果总体标准差已知和未知时采用不同的分布将导致随机不确定度定义的不一致和计算结果不吻合。作者采用蒙特卡罗方法模拟了基于式(7)和基于学生氏 t 分布定义的随机不确定度。发现由式(7)计算的随机不确定度的平均值与式(3)吻合很好,且能够光滑地延伸至 $M=1$ 时的随机不确定度[注意式(7)对 $M=1$ 无效]。然而,基于学生氏 t 分布计算的随机不确定度的平均值只有在子样容量 M 足够大时才趋近于式(3)。当子样容量 M 较小时(如 $M=2,3,4,5$),其结果与式(3)差别很大。而且当 $M=2$ 时,其结果甚至远大于单次测验($M=1$)的随机不确定度,这显然是不合理的。

讨论两个不确定度评估的实际例子,引出两个悖论

例 1: 密西西比河ADCP流量测验

测验次数: 30

平均流量: Q=14,240 m^3/s（流量‘约定真值’）

标准差: σ=223 m^3/s（标准差‘约定真值’）（仅考虑随机误差）

考虑前两次测验:

测验 1 流量: Q_1=14,101 m^3/s

测验 2 流量: Q_2=14,470 m^3/s

两次测验平均流量: Q_{mean}=14,285.5 m^3/s（Q的无偏估计）

样本标准差: s=260.92 m^3/s（σ的有偏估计）

注: ‘约定真值’ = ‘conventional true value’

前两次流量测验不确定度(相对值)

不确定度‘约定真值’（σ已知，*z*-区间半宽） $U_z = z_{95}\frac{\sigma}{\sqrt{n}}$	2.17%
t 不确定度（σ未知用*s*替代，*t*-区间半宽） $U_t = t_{95}\frac{s}{\sqrt{n}}$	16.46%

这个结果显然不可信

ADCP流量测验现场“1.96西格玛”质量保证

ADCP流量测验国际标准ISO:24578:2021(E)规定：

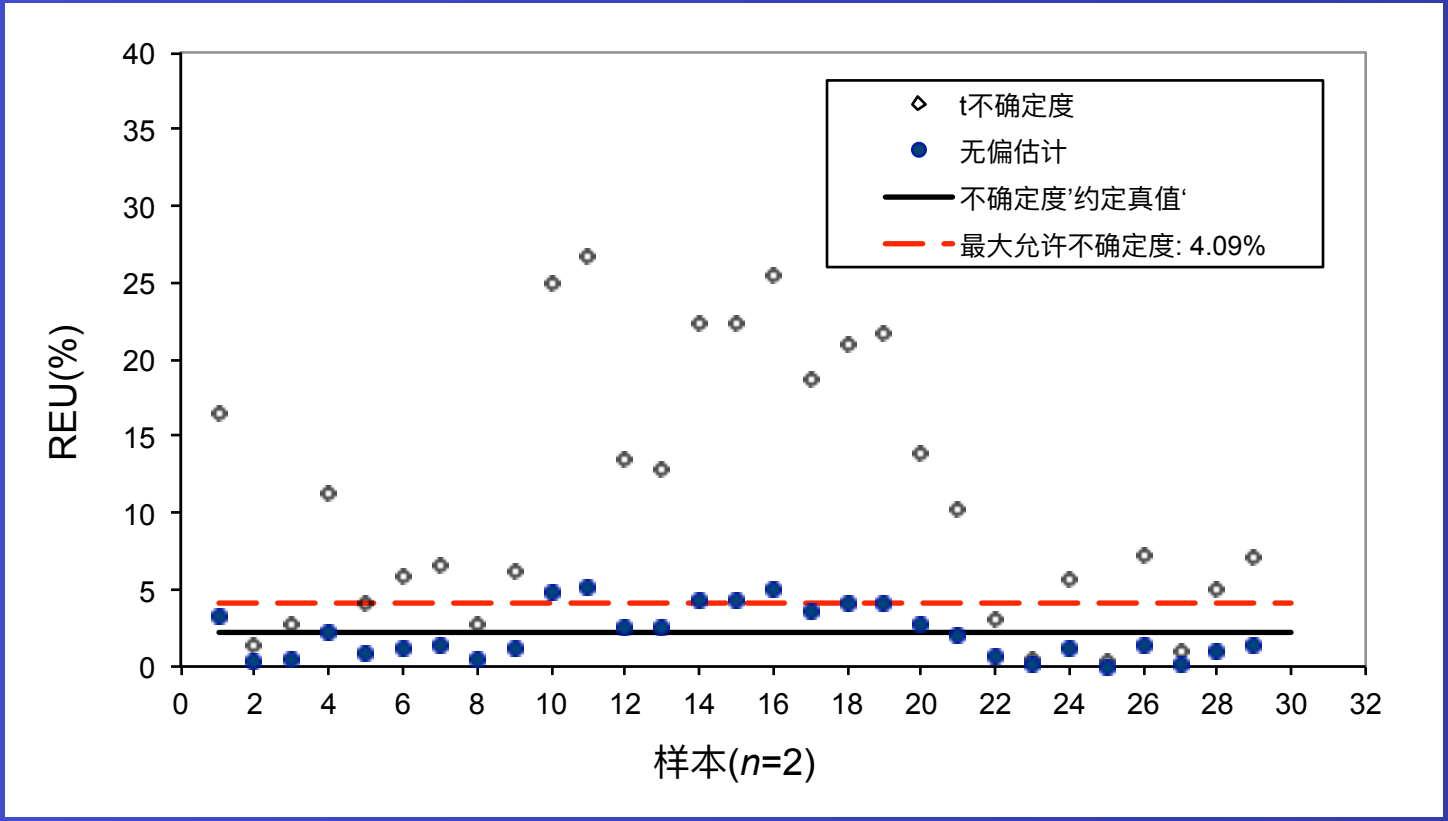
在水流稳定情况下，至少实施往返两次测验（测次n=2，4，6，…），测验总历时不应小于720秒。所有测次流量的平均值作为当前流量值Q_{mean}。如果相对扩展不确定度REU等于或小于最大允许值4.09%，接受当前流量值Q_{mean}作为实测流量，停止测验。否则增加两个测次。所以实际测验中，n=2，4，6，…，直到REU≤4.09%为止。

$$\mathrm{REU}(\text{‘约定真值’}) = z_{95}\frac{\sigma}{Q_{\mathrm{mean}}\sqrt{n}} = 1.96\frac{\sigma}{Q_{\mathrm{mean}}\sqrt{n}} \leq 4.09\%$$

“1.96西格玛”质量保证标准对应于‘标称’保证率95%

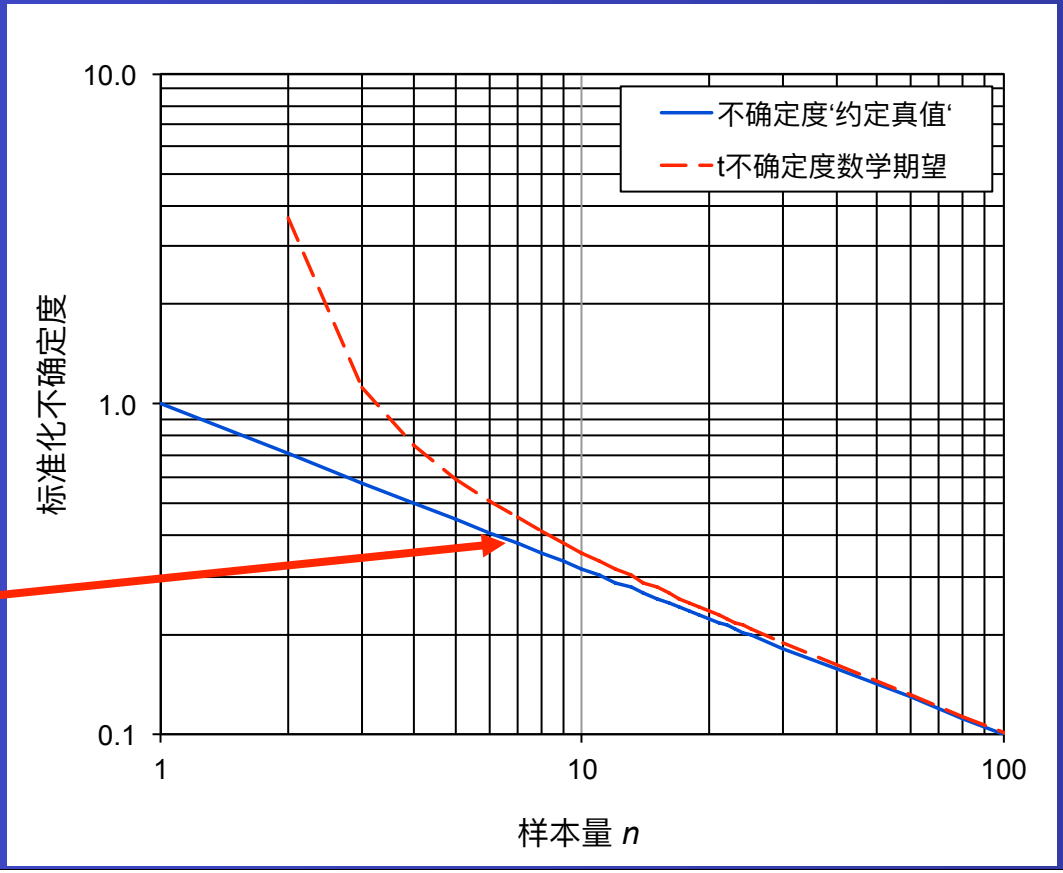
对于密西西比河ADCP测验，按照n=2依此组合得到29个流量值。
因为REU‘约定真值’=2.17%<4.09%，这29个流量值都可以接受作为实测流量，实际保证率为100%。

如果σ未知用s替代，对应于这29个流量值的 *t* 不确定度很大，导致其中21个流量值不被接受：错误拒绝率(false rejection rate)为72%



“不确定度悖论”

对于小样本，*t* 不确定度的数学期望大大偏离不确定度‘约定真值’(也即A类评估与B类评估不一致)



不确定度与样本量的关系遵从-1/2幂定律 -- 物理定律

显然，*t* 不确定度背离物理定律

如何解决“不确定度悖论”?

基于中心极限定理，样本均值量服从正态分布。不确定度‘约定真值’正比于该正态分布的尺度参数。因此，不确定度‘约定真值’的无偏估计等同于尺度参数的无偏估计(统计推断)。即：

$$U_{\text{unbiased}} = z_{95}\frac{s}{c_4\sqrt{n}} \qquad \mathrm{E}(s) = c_4\sigma$$

式中：c_4 =样本标准差偏差修正因子

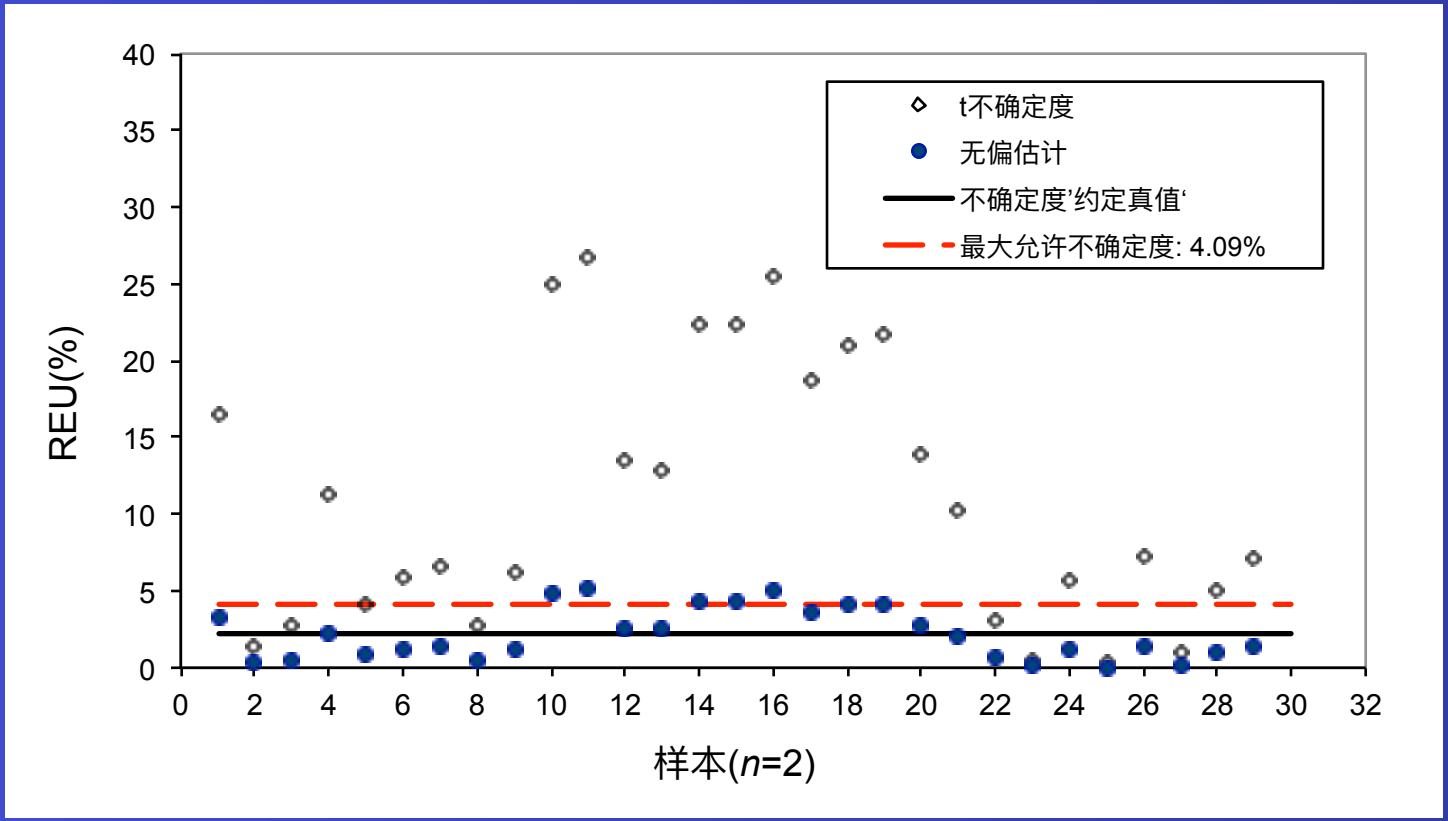
这个**无偏估计方法已经被纳入ADCP流量测验国际标准**
ISO:24578:2021(E)

$$\mathrm{REU}(\text{无偏估计}) = z_{95}\frac{s}{c_4 Q_{\text{mean}}\sqrt{n}} = 1.96\frac{s}{c_4 Q_{\text{mean}}\sqrt{n}} \leq 4.09\%$$

前两次流量测验不确定度(相对值)比较

不确定度‘约定真值’（σ已知，*z*-区间半宽） $U_z = z_{95}\frac{\sigma}{\sqrt{n}}$	2.17%
无偏估计方法（σ未知用*s*替代） $U_{\text{unbiased}} = z_{95}\frac{s}{c_4\sqrt{n}}$	3.18%
t 不确定度（σ未知用*s*替代，*t*-区间半宽） $U_t = t_{95}\frac{s}{\sqrt{n}}$	16.46%

对于密西西比河ADCP测验，无偏估计方法只导致3个流量值不被接受：错误拒绝率为10%，无偏估计方法已经偏于保守。(比较：*t* 不确定度的错误拒绝率为72%)



例 2: “巴利科悖论”

巴利科(Ballico)发现使用《指南》WS-*t* 方法进行不确定度评估的一个反常现象(WS 表示计算有效自由度的Welch-Satterthwaite公式)。巴利科的实验室对一个温度计在两个量程下进行检定:高精度量程 1 mK 和低精度量程10 mK。根据常识, 1 mK 量程不确定度必定小于 10 mK 量程不确定度。然而, 与常识相悖, 1 mK 量程不确定度为 37.39 mK, 大于 10 mK 量程不确定度35.07 mK!

“巴利科悖论”导致无法向客户提供‘检定证书’

为什么会出现“巴利科悖论”?

在WS-*t* 方法中，**扩展不确定度 = 近似 *t* 区间半宽。**

即：

$$U_{p,\mathrm{GUM}} = t_{p,\nu_{\mathrm{eff}}} u_c(y)$$

式中：$t_{p,veff}$ =对应于*p*%置信水平的 *t*-值

v_{eff} =有效自由度（用WS公式计算）

$u_c(y)$=综合标准不确定度（包含A类和B类不确定度分量），根据‘不确定度传播定律’计算

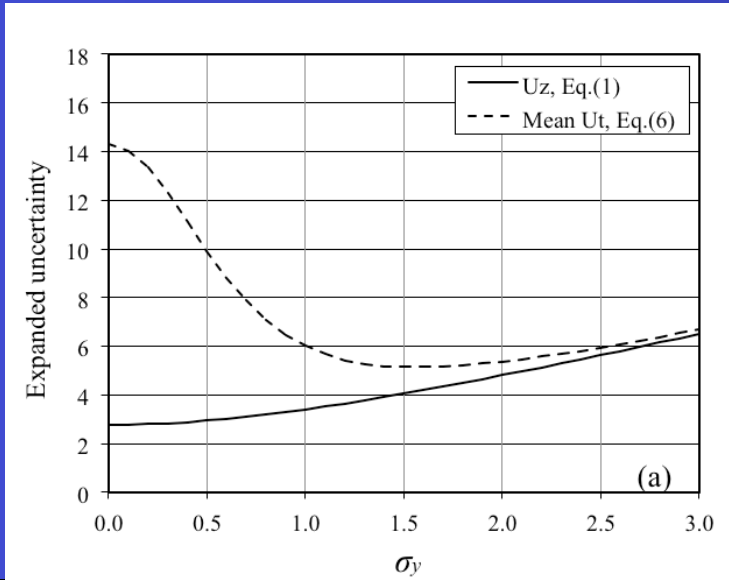
“巴利科悖论”不是偶然的

考虑测量模型：$Z=X+Y$，X、Y服从正态分布。从X中抽取n个值，Y中取一个值，则有：$z=1/n(x_1+x_2+\cdots x_N)+y$。

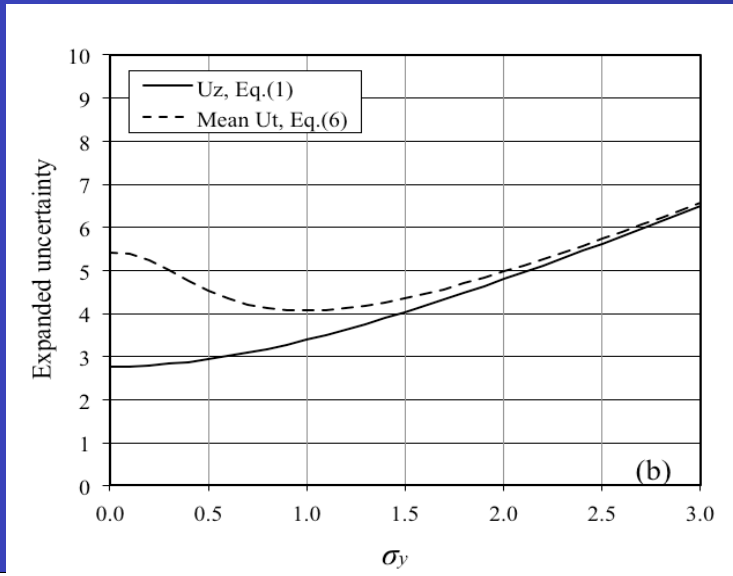
σ_x、σ_y已知，z的**不确定度‘约定真值’为**：$U_z = z_{95}\sqrt{\frac{\sigma_x^2}{n}+\sigma_y^2}$

假定σ_x未知用s_x替代，t **不确定度为**：$U_t = t_{95,\nu_{\mathrm{eff}}}\sqrt{\frac{s_x^2}{n}+\sigma_y^2}$

n=2



n=3



如何解决“巴利科悖论”?

三种方法可以解决“巴利科悖论”，三种方法都基于中心极限定理和无偏估计

方法 1, WS-*z* 方法:

$$U_{p,1} = z_p \frac{u_c(y)}{c_{4,\nu_{\text{eff}}}}$$

方法 2:

$$U_{p,2} = z_p u'_c(y)$$

所有A类不确定度分量用c_4 修正

方法 3:

z-based MCM（蒙特卡洛方法），使用 scaled and shifted *z*-分布（即正态分布）

“巴利科悖论”的三种解决方法

“巴利科悖论”的三种解决方法及与《指南》WS-*t* 方法的比较

Method	Expanded uncertainty (mK)	
	1 mK range	10 mK range
GUM’s WS-*t* approach (Ballico 2000), Eq. (2)	37.39	35.07
First alternative approach (WS-*z* approach, (Huang 2016)), Eq. (4)	25.75	29.34
Second alternative approach (Huang 2018), Eq. (8)	25.95	29.95
z-based MCM (Huang 2019)	25.99	29.94

发现‘*t*-转换扭曲’

- **造成“不确定度悖论”和“巴利科悖论”的根源是什么？**
- 这个问题困惑了我整整9年（2006至2015），直到2015年2月读到Harrell一篇博文中的一句话：“Playing with transformations distorts every part of statistical inference…（使用[变量]转换会扭曲统计推断的每个部分）”，茅塞顿开。
- 问题的根源是：原始样本空间 Ω（ε，s）通过‘*t*-转换’变成了Ω（t，s）。基于*t*-分布的推断是在‘*t*-转换’的样本空间 Ω（t，s）进行的。然而，‘*t*-转换’扭曲了统计推断。

比如将柿子做成柿子饼是一种“转换”。根据柿子饼的形状 --“扭曲”的形状 -- 不可能正确推断柿子的形状。



https://baike.sogou.com/v133489.htm

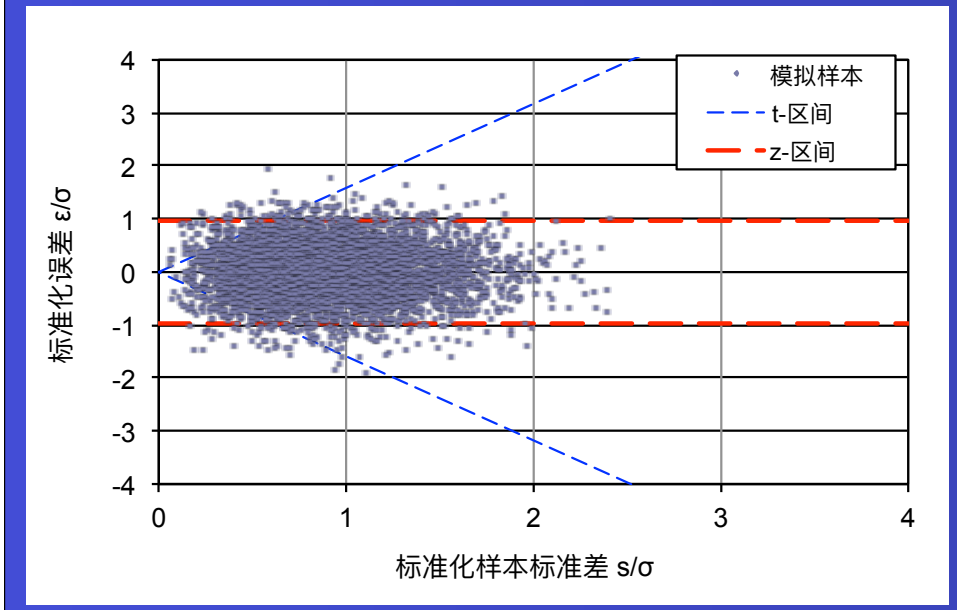


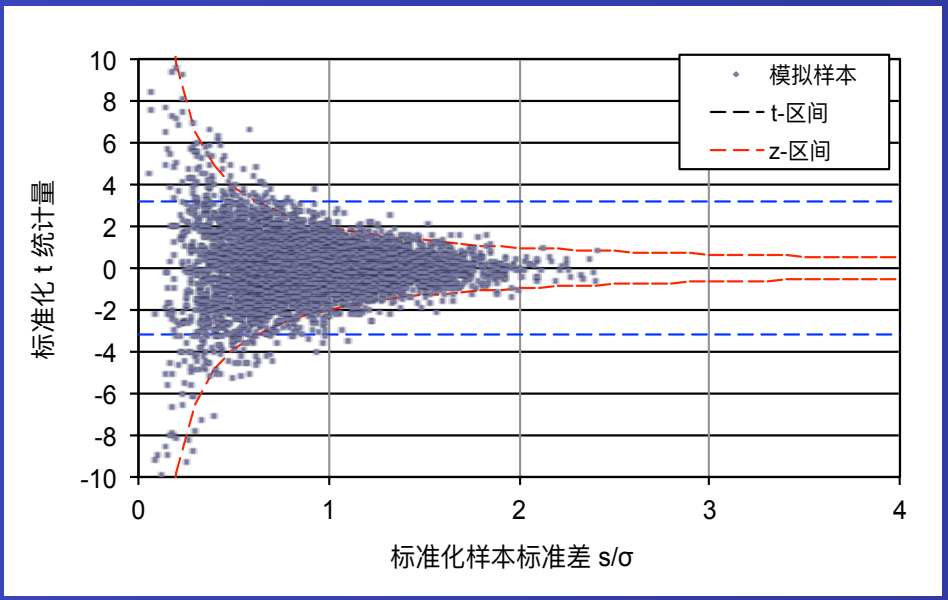
‘*t*-转换扭曲’

下图显示仿真模拟的样本误差 ε -- 样本标准差*s*数据（用σ标准化）：*n*=4

原始物理平面的数据：*z*-区间边界为两条平行线，*t*-区间扇形区域边界为两条斜线

t-转换平面的数据：*z*-区间和数据发生了扭曲，*t*-区间边界变为两条平行线





t-区间与z-区间的比较(原始物理平面)

z-区间(带形区域):

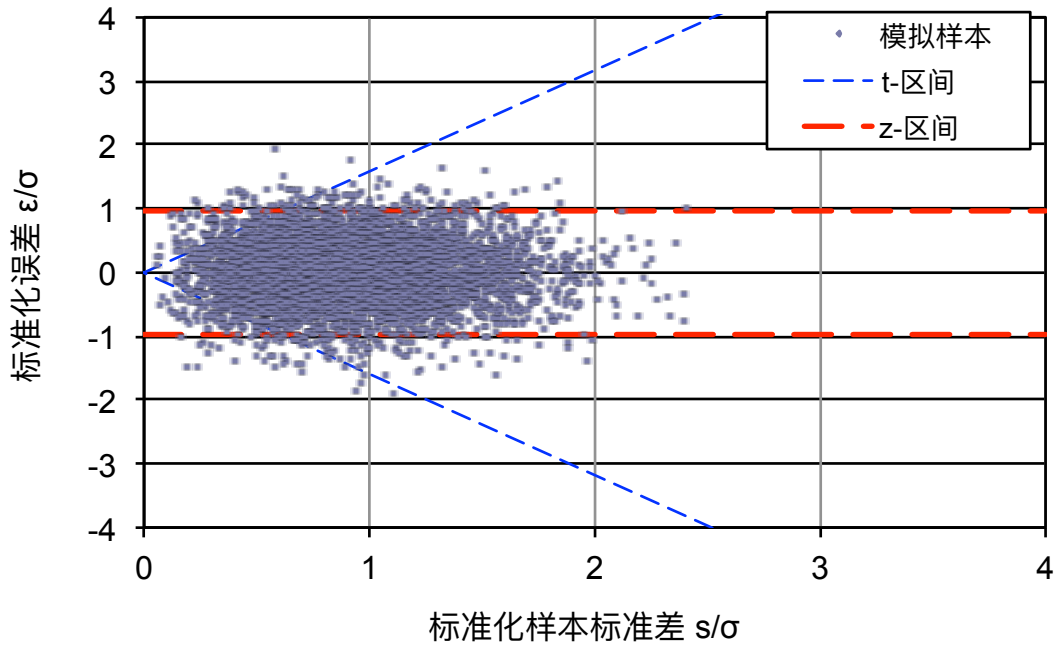
$$(-\frac{z_{95}}{\sqrt{n}} \le \frac{\varepsilon}{\sigma} \le +\frac{z_{95}}{\sqrt{n}})$$

t-区间(扇形区域):

$$(-\frac{t_{95}s}{\sigma\sqrt{n}} \le \frac{\varepsilon}{\sigma} \le +\frac{t_{95}s}{\sigma\sqrt{n}})$$

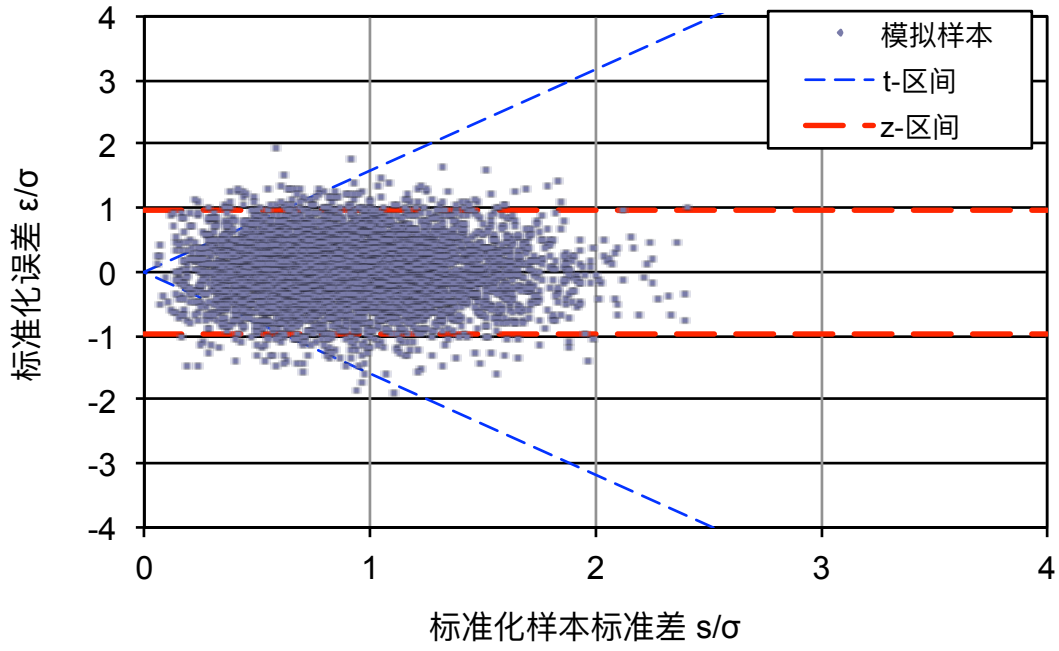
t-值=*z*-值时，z-区间和*t*-区间的“覆盖概率”相同。例如：*p*=95%。但是，它们分别“覆盖”不同的数据(样本)。

也就是说，*t*-区间的“内容”不同于z-区间的“内容”。



教科书对*t*-值的解释是错误的

在统计学教科书中，*t*-值被认为与*z*-值具有相同的意义: 量度样本均值与总体均值的距离；对于相同的覆盖概率，*t*-分布的尾部较厚重,所以*t*-值大于*z*-值，意味着*t*-分布更容易产生远离总体均值的值。



为什么这种解释是错误的？

（1）*z*-值是*z*-区间边界水平线的截距因子，*t*-值是*t*-区间边界斜线的斜率因子；水平线和斜线不可能重合。所以即使样本量很大时*t*-值接近*z*-值，*t*-值的数学意义依然与*z*-值不同。

（2）*t*-分布的尾部是由*t*-区间扇形区域以外的数据构成的，其中包含大量误差很小的数据，即它们对应的样本均值实际上与总体平均值的距离很小。

关于学生氏*t*-分布的几点澄清

- *t* 不是现实世界的物理量；它没有任何物理意义：*t* = 噪声 ε /噪声*s* 是什么意思？
- 学生氏 *t*-分布不是‘误差 ε ’的分布；它是一个扭曲的 *z*-分布（由于 *t*-转换扭曲）。
- Scaled and shifted ***t*-分布不是样本均值量的抽样分布。**
- **样本均值量的抽样分布**是 scaled and shifted ***z*-分布（即正态分布）**，无论**尺度参数** σ 已知还是未知。如果 σ 未知，**可以取其‘点估计’值，无论样本量大小。**

关于学生氏*t*-分布的几点澄清(续)

- **学生氏 *t*-分布的数学推导没有问题。但是，由于 *t*-转换扭曲，基于学生氏 *t*-分布进行统计推断是错误的。这是方法论的问题。**
- 贝叶斯方法采用1/σ先验得到 scaled and shifted *t*-分布完全是“凑巧”。采用其它先验，例如$1/\sigma^2$，将得到不同的分布。有贝叶斯学者认为1/σ先验不符合正态分布参数的物理意义。

结论和建议

- **“不确定度悖论”和“巴利科悖论”是基于学生氏 *t*-分布推断的反例，足以证伪**
- **学生氏*t*-分布误导了小样本统计推断**
- **小样本统计推断应基于中心极限定理在原始物理平面进行**
- 无偏估计方法可以给出不确定度的合理估计，并且符合-1/2幂定律
- **《指南》修正版应放弃置信区间，采用无偏估计方法**
- **应重新检查其它基于*t*-分布的推断方法**

相关文献

关于*t*-转换扭曲、无偏估计方法、“不确定度悖论”

Huang, H. (2018). Uncertainty estimation with a small number of measurements, Part I: new insights on the *t*-interval method and its limitations. *Measurement Science and Technology*, 29(1), 015004, https://doi.org/10.1088/1361-6501/aa96c7

Huang, H. (2018). Uncertainty estimation with a small number of measurements, Part II: a redefinition of uncertainty and an estimator method. *Measurement Science and Technology*, 29(1), 015005, https://doi.org/10.1088/1361-6501/aa96d8

Huang, H. (2018). More on the *t*-interval method and mean-unbiased estimator for measurement uncertainty estimation. *Cal Lab the International Journal of Metrology* 25(2), 24-33.

Huang, H. (2018). A unified theory of measurement errors and uncertainties. *Measurement Science and Technology,* 29(12), 125003, https://doi.org/10.1088/1361-6501/aae50f

关于“巴利科悖论”

Ballico, M. (2000). Limitations of the Welch-Satterthwaite approximation for measurement uncertainty calculations. *Metrologia* 37:61-64

Huang, H. (2016). On the Welch-Satterthwaite formula for uncertainty estimation: a paradox and its resolution. *Cal Lab the International Journal of Metrology*, 23(4), 20-28.

关于scaled and shifted *t*-分布

Huang, H. (2019). Why the scaled and shifted *t*-distribution should not be used in the Monte Carlo method for estimating measurement uncertainty? *Measurement*, 136, 282-288, https://doi.org/10.1016/j.measurement.2018.12.089

关于ADCP流量测验现场“1.96西格玛”质量保证

Huang, H. (2015). Statistical quality control of streamflow measurements with moving-boat acoustic Doppler current profilers. *Journal of Hydraulic Research,* 53(6), 820-827, https://doi.org/10.1080/00221686.2015.1074947

ISO:24578:2021(E), Hydrometry — Acoustic Doppler profiler — Method and application for measurement of flow in open channels from a moving boat, first edition, 2021-3.

谢谢！