在动物实验数据中剔除坏值的建议*

周 文 扬

(中国科学院西北高原生物研究所)

粗差 (gross error) 可定义为：明显歪曲实验结果的误差。含有粗差的实验值称为坏值(anomalous)、亦称异常值或反常值。它是由于实验者主观上的疏忽，或客观条件的异常变化造成的。如实验中的读错、记错、操作不当、仪器有毛病以及其它许多我们估计不到的意外情况。总之，造成粗差的情况，在所有实验中都可能发生，而对某一次实验则无法预知，往往不出

* 本文承王祖望同志提供图 1 的实验数据并提出宝贵意见，特此致谢。

现，有时在实验进行中可以发现并排除，有时则在实验当时未能发现甚至不可能发现，于是在结果中形成坏值。

数理统计原理是建立在变异量的概率分布基础上的。如果存在这样的量，它的数值超过了实验随机误差的范围，在置信区间已规定的条件下，可以认为它不属于这一概率分布，应将它除去。粗差便是这种情形，这就是为什么在统计学中应当发现粗差并予以排除的根据。由于粗差而造成的坏值，往往使研究样品的概率分布发生变化，甚至导致错误的结论。比如，使统计误差大于实验实际的随机误差，使差异显著性测定与它本来的客观规律不一致，使相关回归关系与实际偏离等。

过去工作中，确实出现过这种情况。例如，在青海省草原灭鼠中（高原鼠兔 *Ochotona curzoniae*）出现磷化锌灭鼠效果逐年降低。从 1965 年到 1970 年，我们曾进行了新灭鼠药甘氟的筛选及大面积群众灭鼠试验，并始终以磷化锌作对照。据历年试验统计，磷化锌平均灭鼠效果为 93.2±1.8（单位：百分率。下同），甘氟历年在不同地区的灭鼠效果为：98.7、55.0、97.6、97.2、96.4、94.8、96.4，总的灭鼠效果为 90.9±6.0。经 t 值检验，与磷化锌无显著差异

$(t = 0.34,\ t_{0.05} = 2.447,\ P > 0.05)$。

问题在于，55.0一数系 1968 年冬在青海省河南蒙旗县 1,200 亩灭鼠试验中所得，当年由于该地气候异常，气温偏高，土表未冻实。鼠易觅食草根，不喜食毒饵，致使灭鼠效果降低，并非甘氟本身的问题，故此数据值得怀疑。经统计计算，该数据确系坏值，应予剔除（计算见表 4）。重新统计结果表明，平均灭鼠效果应为 96.8±1.4，再经 t 测，与磷化锌差异极其显著

$(t = 6.0,\ t_{0.01} = 4.032,\ P < 0.01)$，

说明甘氟灭鼠效果显著优于磷化锌。这与实际情况较为符合，两种药物虽均为高效杀鼠剂，但磷化锌由于连年使用，使鼠产生了忌避性。观察亦表明，鼠兔对甘氟毒饵的喜食程度高于磷化锌毒饵。

在实验当时发现明显误差，可随时纠正或排除，不必列入结果记录中。但有时整个实验完成之后，也无法断定是否存在坏值，即便有的数据看起来偏离程度相当大，也决不能凭主观意愿取舍，这就必须借助数理统计方法判断。同时，还应当指出：即便发现坏值，除非确实有因可查，否则在剔除异常数据时必须持慎重态度，以免将具有某些特殊规律的数据忽略掉。数理统计只为我们提供了判别异常数据的方法，到底是否应该剔除还取决于实践的本身。

方 法 简 介

统计学中常见的几种判别方法有 3σ 准则、肖维勒（Chauvenet）准则、格拉布斯（Grubbs）准则、t 检验准则、狄克逊（Dixon）准则等[1]。

1. 格拉布斯准则

若某组实验数据为 $x_1, \cdots, x_n$，其算术平均值为 $\bar{x}$，离均差

$$V_i = x_i - \bar{x}\ (i = 1, \cdots, n),$$

标准差为 σ。其中离均差的绝对值最大的数值以 $|V_d| = |x_d - \bar{x}|$ 表示，x_d 为实验数据中偏离平均值最大的一个。当

$$|V_d| > \lambda(\alpha, n) \cdot \sigma$$

时，则认为 x_d 是坏值，可予剔除。$\lambda(\alpha, n)$ 值列于表 1，其中 α 称显著性水平，取值（0.01 或 0.05）视实验所取置信概率而定。坏值剔除后，$\bar{x}$ 及 σ 等值应重新计算。

2. t 检验准则

先除去可疑值 x_d 计算平均值 $\bar{x}$ 和标准差 σ：

$$\bar{x} = \frac{1}{n-1} \sum_{\substack{i=1,\\ i \neq d}}^{n}$$

$$\sigma = \sqrt{\frac{1}{n-2} \sum_{\substack{i=1\\ i \neq d}}^{n} V_i^2}$$

$\neq d$ 表示计算中不包括 x_d。当

$$|V_d| > K(\alpha, n) \cdot \sigma$$

时，可剔除坏值 x_d。去掉坏值后，$\bar{x}$ 及 σ 无需再

1) 张世英，刘智敏 1977 测量实践的数据处理，366—369、484—489。科学出版社。
Dixon, W. J. 1950 Analysis of extreme values, Ann. of Math. Stat. 21.

表 1　$\lambda(\alpha, n)$ 值表

$n\backslash\alpha$	0.05	0.01	$n\backslash\alpha$	0.05	0.01
3	1.15	1.16	17	2.48	2.78
4	1.46	1.49	18	2.50	2.82
5	1.67	1.75	19	2.53	2.85
6	1.82	1.94	20	2.56	2.88
7	1.94	2.10	21	2.58	2.91
8	2.03	2.22	22	2.60	2.94
9	2.11	2.32	23	2.62	2.96
10	2.18	2.41	24	2.64	2.99
11	2.23	2.48	25	2.66	3.01
12	2.28	2.55	30	2.74	3.10
13	2.33	2.61	35	2.81	3.18
14	2.37	2.66	40	2.87	3.24
15	2.41	2.70	50	2.96	3.34
16	2.44	2.75	100	3.17	3.59

表 2　t 检验 $k(\alpha, n)$ 值表

$n\backslash\alpha$	0.01	0.05	$n\backslash\alpha$	0.01	0.05	$n\backslash\alpha$	0.01	0.05
4	11.46	4.97	13	3.23	2.29	22	2.91	2.14
5	6.53	3.56	14	3.17	2.26	23	2.90	2.13
6	5.04	3.04	15	3.12	2.24	24	2.88	2.12
7	4.36	2.78	16	3.08	2.22	25	2.86	2.11
8	3.96	2.62	17	3.04	2.20	26	2.85	2.10
9	3.71	2.51	18	3.01	2.18	27	2.84	2.10
10	3.54	2.43	19	3.00	2.17	28	2.83	2.09
11	3.41	2.37	20	2.95	2.16	29	2.82	2.09
12	3.31	2.33	21	2.93	2.15	30	2.81	2.08

表 3　狄克逊系数 $f(\alpha, n)$ 与 f_0 计算公式表

n	$f(\alpha, n)$		f_0 计算公式	
	$\alpha = 0.01$	$\alpha = 0.05$	x_1 可疑时	x_n 可疑时
3	0.988	0.941	$\dfrac{x_2 - x_1}{x_n - x_1}$	$\dfrac{x_n - x_{n-1}}{x_n - x_1}$
4	0.889	0.765		
5	0.780	0.642		
6	0.698	0.560		
7	0.637	0.507		
8	0.683	0.554	$\dfrac{x_2 - x_1}{x_{n-1} - x_1}$	$\dfrac{x_n - x_{n-1}}{x_n - x_2}$
9	0.635	0.512		
10	0.597	0.477		
11	0.679	0.576	$\dfrac{x_3 - x_1}{x_{n-1} - x_1}$	$\dfrac{x_n - x_{n-2}}{x_n - x_2}$
12	0.642	0.546		
13	0.615	0.521		
14	0.641	0.546	$\dfrac{x_3 - x_1}{x_{n-2} - x_1}$	$\dfrac{x_n - x_{n-2}}{x_n - x_3}$
15	0.616	0.525		
16	0.595	0.507		
17	0.577	0.490		
18	0.561	0.475		
19	0.547	0.462		
20	0.535	0.450		
21	0.524	0.440		
22	0.514	0.430		
23	0.505	0.421		
24	0.497	0.413		
25	0.489	0.406		

算。$K(\alpha, n)$ 值列于表 2。

3. 狄克逊准则

这是一种用极差比的方法，得到简化而严密的结果，并且根据数据的多少采用不同的极差比以提高判别准确性。公式及系数列于表 3。

判断时不需计算平均值和标准差，但要先将 n 个数据按大小重新排列为

$$x_1 \leqslant x_2 \leqslant \cdots\cdots \leqslant x_n,$$

再由表中选择相应公式计算 f_0 值。当

$$f_0 > f(\alpha, n)$$

时，x_1（或 x_n）为坏值。x_1、x_2、x_3 分别为前三位小的数据，x_n、x_{n-1}、x_{n-2} 分别为后三位大的数据，参加运算的仅这几个数。所以，数据较多时，实际只需挑出这几个数据即可。

应　用　举　例

结合动物试验，对各判断准则的实际使用试举数例：

1. 对甘氟灭鼠效果数据的检验。数据处理方法及结果列于表 4：

2. 一次动物热值测定中 15 个实验数据的检验，方法及结果列于表 5：

3. 同一生境中，四块样地用铗日法捕打小家鼠的捕鼠率分别为(%)：15、15、10、4，是否属同一概率分布？

其中数值 4 颇有怀疑，经计算：$\bar{x} = 11.0$，$\sigma = 5.2$ $|x_d - \bar{x}| = 7$，按格拉布斯准则判断，$n = 4$ 时，

$$\lambda(0.05, 4) = 1.46,$$

$$\lambda(0.05, 4) \cdot \sigma = 7.6,$$

$$|x_d - \bar{x}| < \lambda(0.05, 4) \cdot \sigma,$$

表 4　甘氟灭限试验中坏值的剔除

n	灭效(%)	计　算　1	计　算　2	格拉布斯准则	t 检验准则	狄克逊准则
	x_i	$x_i-\bar{x}_{(1)}$	$x_i-\bar{x}_{(2)}$			
1	55.0	−35.9	(−41.8)*	计算 1：按全部数据计算 $\sigma_{(1)}=15.9$，查表 1：n 为 7 时 $\lambda(0.05,7)=1.94$ $\lambda(0.05,7)\cdot\sigma_{(1)}=30.8$ 35.9>30.8 55.0 一数为坏值。 计算 2：坏值剔除后重算 $\sigma_{(2)}=1.7$，$n=6$ 时 $\lambda(0.05,6)=1.82$ $\lambda(0.05,6)\cdot\sigma_{(2)}=3.09$ 与 $\lvert x_i-\bar{x}_{(2)}\rvert$ 比较，再无坏值。	先将可疑值55.0剔除计算 $\sigma_{(2)}=1.7$ 查表 2，$n=7$ 时 $K(0.05,7)=2.78$ $K(0.05,7)\cdot\sigma_{(2)}=4.7$ 41.8>4.7 55.0 一数为坏值。 再与其余 $\lvert x_i-\bar{x}_{(2)}\rvert$ 值比较，无坏值。	查表 3，$n=7$，x_1 可疑时用公式计算 $f_0=\frac{x_2-x_1}{x_n-x_1}=\frac{94.8-55.0}{98.7-55.0}=0.91$ 查表 $f(0.05,7)=0.507$ $f_0>f(0.05,7)$ 55.0 一数为坏值。
2	94.8	3.9	−2.0			
3	96.4	5.5	−0.4			
4	96.4	5.5	−0.4			
5	97.2	6.3	0.4			
6	97.6	6.7	0.8			
7	98.7	7.8	1.9			
$\bar{x}$		$\bar{x}_{(1)}=90.9$	$\bar{x}_{(2)}=96.8$			
σ		$\sigma_{(1)}=15.9$	$\sigma_{(2)}=1.7$			

* 括弧中数据不参加计算，仅在 t 检验准则中作比较用(下表同)。

表 5　一次动物热值测定中坏值的剔除

n	x_i	计算 1	计算 2	计算 3	格拉布斯准则	t 检验准则
		$x_i-\bar{x}_{(1)}$	$x_i-\bar{x}_{(2)}$	$x_i-\bar{x}_{(3)}$		
1	36.47	−0.02	−0.01	0.00	计算 1：按全部数据计算 $\bar{x}_{(1)}$ 及 $\sigma_{(1)}$，$n=15$ 查表 1， $\lambda(0.05,15)=2.41$ $\lambda(0.05,15)\cdot\sigma_{(1)}=0.188$ $\lvert x_7-\bar{x}_{(1)}\rvert=0.20$ 0.20>0.188 36.69 为坏值。 计算 2：余下 14 个数据计算 $\bar{x}_{(2)}$ 及 $\sigma_{(2)}$， $n=14$　$\lambda(0.05,14)=2.37$ $\lambda(0.05,14)\cdot\sigma_{(2)}=0.114$ $\lvert x_{10}-\bar{x}_{(2)}\rvert=0.12$ 0.12 > 0.114 36.60 也是坏值。 计算 3：余下 13 个数据计算 $\bar{x}_{(3)}$ 及 $\sigma_{(3)}$， $n=13$　$\lambda(0.05,13)=2.33$ $\lambda(0.05,13)\cdot\sigma_{(3)}=0.077$ 对照 $\lvert x_i-\bar{x}_{(3)}\rvert$ 各值再无坏值。	先将可疑值 x_7、x_{10} 剔除，对余下 13 个数据计算 $\bar{x}_{(3)}$ 及 $\sigma_{(3)}$ 查表 2， $n=15$　$K(0.05,15)=2.24$ $K(0.05,15)\cdot\sigma_{(3)}=0.074$ 经检验： $\lvert x_7-\bar{x}_{(3)}\rvert=0.22$ 0.22>0.074， $\lvert x_{10}-\bar{x}_{(3)}\rvert=0.13$ 0.13>0.074 36.69 及 36.60 确系坏值。
2	36.40	−0.09	−0.08	−0.07		
3	36.49	0.00	0.01	0.02		
4	36.53	0.04	0.05	0.06		
5	36.46	−0.03	−0.02	−0.01		
6	36.44	−0.05	−0.04	−0.03		
7	36.69	0.20		(0.22)		
8	36.43	−0.06	−0.05	−0.04		
9	36.49	0.00	0.01	0.02		
10	36.60	0.11	0.12	(0.13)		
11	36.47	−0.02	−0.01	0.00		
12	36.50	0.01	0.02	0.03		
13	36.48	−0.01	0.00	0.01		
14	36.46	−0.03	−0.02	−0.01		
15	36.45	−0.04	−0.03	−0.02		
$\bar{x}$		$\bar{x}_{(1)}=36.49$	$\bar{x}_{(2)}=36.48$	$\bar{x}_{(3)}=36.47$		
σ		$\sigma_{(1)}=0.072$	$\sigma_{(2)}=0.048$	$\sigma_{(3)}=0.033$		
注		包括全部数据	剔除坏值 36.69 后计算	剔除坏值 36.69 及 36.60 后计算		

狄　克　逊　准　则
先将15个数据由小到大按序排列(中间数值可不必写出)： $x_1\ x_2\ x_3\ \cdots x_{n-2}\ x_{n-1}\ x_n$ 36.40　36.43　36.44……36.53 36.60　36.69 查表 3，x_n 可疑、$n=15$ 时用公式 $f_0=\frac{x_n-x_{n-2}}{x_n-x_3}=0.640$ $f(0.05,15)=0.525$ $f_0>f(0.05,15)$ 36.69 为坏值。 余下 14 个数据后，原 x_{n-1} 一数递补为 x_n，若仍属可疑，$n=14$ 时依然用上述公式： $f_0=\frac{x_n-x_{n-2}}{x_n-x_3}=0.625$ (式中 x_{n-2} 已不是 36.53 一数，而应由表中 36.50 一数递补) $f(0.05,14)=0.546$ $f_0>f(0.05,14)$ 36.60 也是坏值。

说明此数并非坏值，本样地数据均在随机误差范围之内。

4. 某课题对青海省海西地区的高原鼠兔和中华鼢鼠（*Myospalax fontanierii*）进行了不同温度条件下耗氧量的测定，获得大量实验数据。这些数据中包含了某些个体的特殊差异和实验粗差，但实验当时往往不易发现。因此，在进行统计分析的过程中，发现一些数据异常，又不能主观决定取舍，给统计造成一定困难。若根据统计原理，事先剔除实验数据中的坏值，再进一步进行数学分析，所得结论，可能更符合客观规律。

按照动物能量代谢规律，环境温度在一定范围，哺乳动物体内与环境温度的热量交换趋于平衡，此时动物体耗氧量最低（就基础代谢而言）。温度低于或高于此范围，耗氧量都会增加，从而增加产热或通过体表水分蒸发加强散热，以维持体温恒定。环境温度—耗氧量曲线应呈单谷形，只有一个最低点，其对应的温度值称临界温度（图 1 曲线 III），中华鼢鼠的理论临界温度平均值为 25.01℃，高原鼠兔为 26.68℃[1]。但是，某些实验却出现反常，曲线最低点不止一处（图 1 曲线 I），经统计检查，确系由坏值造成，使该组的平均值提高。将坏值剔除后重新统计作图，曲线继续向临界温度下降，更加实地反映了客观情况（见图 1 曲线 II）。

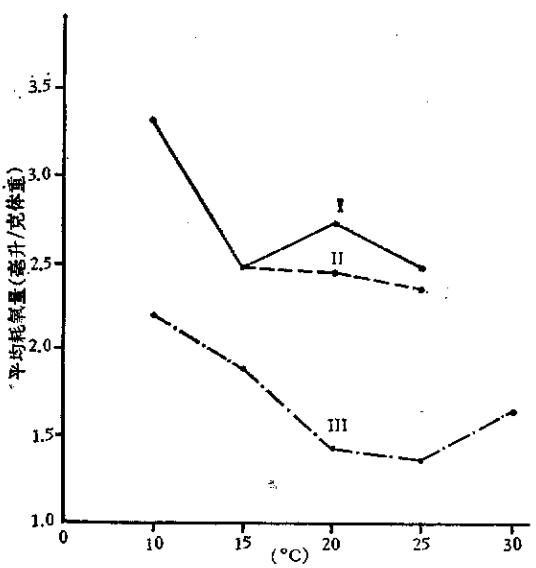

图 1　不同温度下两种鼠的耗氧量
═══高原鼠兔　—·—中华鼢鼠

讨　论

1. 从例中可看出，若坏值不止一个，则以 t 值检验准则较适宜，仅一次计算就可剔除所有坏值，并得到正确的平均值和标准差。而数据较多且无需计算平均值和标准差时，狄克逊准则只需几次简单计算就能将坏值剔除。此外，几种方法虽均有各自的理论依据及精确的数学推导，但统计结果有时并不一致，遇到这种情况更要慎重，一般不要轻易剔除。

2. 坏值及可疑值的离均差与均数之比（即相对变异量 $V=(x_d-\bar{x})/\bar{x}$）、反映出该值与均数的偏离程度。将几例的 V 值进行比较：例 3 中，$V=0.64$，并无坏值出现；例 1 中，$V=0.39$，剔除一个坏值；例 2 中，$V<0.006$，坏值却不只一个。这说明，与均数偏离程度大的数据，较易受到怀疑，但并不一定是坏值。若不经统计方法检验，仅凭主观感觉决定数据取舍往往可能导致错误判断。一般地说，在精密度较高及数据较多的实验中，容易让坏值保存下来，影响结果的准确性；反之，在数据较少或较粗糙的实验中，容易不适当地取掉某些偏离程度大的数据，造成统计结果不能如实反映实验数据的真正概率分布。

1）两种鼠均采自海西州石陂地区。