中华人民共和国国家标准

UDC 519.28

GB 3361—82

# 数据的统计处理和解释 在成对观测值情形下两个均值的比较

Statistical interpretation of data Comparison of two means in the case of paired observations

本标准规定了成对观测值之差的总体均值与零或者其它预先指定的值相比较的方法。

对两个具有某种特性的观测值$X_i$和$Y_i$，如果是在如下情形获得，则称它们是成对观测值：

**a.** 取自同一总体的同一个体，但观测条件不同（例如：同一产品的两种不同分析方法结果的比较）。

**b.** 两个不同的个体，除了检验所涉及的系统差异外，其它所有方面都相似（例如：播种两种不同品种的种子的两块相邻土地的产量）。

必须注意，在情形b中,检验的功效依赖于如下的假设是否正确：在成对的个体之间，除了所检验的系统差异外，不存在其它的系统差异。

本标准系参考国际标准ISO 3301《数据的统计解释——在成对观测值情形下两个均值的比较》（1975年第一版）制订的。

## 1 应用的范围

这个方法可用来确定两种处理间的差异。在这种情形,$X_i$是第一种处理的第i个观测值,$Y_i$是第二种处理的第i个观测值,这两个观测结果系列是不独立的。术语“处理”应该理解为广义的。例如：所比较的两种处理可以是两种检验方法，两台仪器或者两个实验室，以便发现两种处理之间的可能的系统误差。用同样的试验材料相继进行的两种处理可能相互影响，获得的值与次序有关。优良的试验设计应该能消除这种偏倚。另外，也可用于仅有一个处理的情形，它的效应可以与无处理时相比较，这种比较的目的是确定该处理的效应。

## 2 应用的条件

如果满足下列两个条件，则这个方法能够有效地应用：

**a.** 差$d_i = X_i - Y_i$的系列看作独立随机变量系列；

**b.** $d_i$的分布是正态分布或近似正态分布。

如果$d_i$的分布偏离正态，则当样本大小充分大时，所述的方法仍然有效；偏离正态越大，需要的样本大小也越大。然后，即使在特殊的情形，样本大小为100,对于大部分的实际应用是足够的。

国家标准局1982－12－30发布　　　　1984－01－01实施

## 3 计算公式表

所研究的问题……

试验条件……

| 统计项目 | 计算 |
|---|---|
| 样本大小：<br>$n =$<br>观测值的和：<br>$\sum_i X_i = \sum_i Y_i =$<br>差的和：<br>$\sum_i d_i =$<br>差的平方和：<br>$\sum_i d_i^2 =$<br>给定值：<br>$d_0 =$<br>自由度：<br>$\nu = n - 1 =$<br>显著性水平：<br>$\alpha =$ | $\bar{d} = \frac{1}{n}(\sum_i X_i - \sum_i Y_i)$<br>$= \frac{1}{n}\sum_i d_i =$<br>$S_d^2 = \frac{1}{n-1}[\sum_i d_i^2 - \frac{1}{n}(\sum_i d_i)^2] =$<br>$\hat{\sigma}_d = \sqrt{S_d^2}$<br>$A_1 = [t_{1-\alpha}(\nu)/\sqrt{n}]\ \hat{\sigma}_d =$<br>$A_2 = [t_{1-\alpha/2}(\nu)/\sqrt{n}]\ \hat{\sigma}_d =$ |

结果

双侧情形：

若 $|\bar{d} - d_0| > A_2$

则拒绝差的总体均值 $D$ 等于 $d_0$ 的假设。

单侧情形：

a. 若 $\bar{d} < d_0 - A_1$

则拒绝差的总体均值 $D$ 至少等于 $d_0$ 的假设。

b. 若 $\bar{d} > d_0 + A_1$

则拒绝差的总体均值 $D$ 至多等于 $d_0$ 的假设。

注：$t_{1-\alpha}(\nu)$ 是自由度为 $\nu$ 的 $t$ 变量的 $(1-\alpha)$ 分位数。

$t_{1-\alpha}(\nu)/\sqrt{n}$ 的值在表 1 中给出。

表1 比值$t_{1-\alpha}(\nu)/\sqrt{n}$ ($\nu=n-1$)

| $\nu=n-1$ | 双侧情形 | | 单侧情形 | |
|---|---|---|---|---|
| | $t_{0.975}/\sqrt{n}$ | $t_{0.995}/\sqrt{n}$ | $t_{0.95}/\sqrt{n}$ | $t_{0.99}/\sqrt{n}$ |
| 1 | 8.985 | 45.013 | 4.465 | 22.501 |
| 2 | 2.484 | 5.730 | 1.686 | 4.021 |
| 3 | 1.591 | 2.920 | 1.177 | 2.270 |
| 4 | 1.242 | 2.059 | 0.953 | 1.676 |
| 5 | 1.049 | 1.646 | 0.823 | 1.374 |
| 6 | 0.925 | 1.401 | 0.734 | 1.188 |
| 7 | 0.836 | 1.237 | 0.670 | 1.060 |
| 8 | 0.769 | 1.118 | 0.620 | 0.966 |
| 9 | 0.715 | 1.028 | 0.580 | 0.892 |
| 10 | 0.672 | 0.956 | 0.546 | 0.833 |
| 11 | 0.635 | 0.897 | 0.518 | 0.785 |
| 12 | 0.604 | 0.847 | 0.494 | 0.744 |
| 13 | 0.577 | 0.805 | 0.473 | 0.708 |
| 14 | 0.554 | 0.769 | 0.455 | 0.678 |
| 15 | 0.533 | 0.737 | 0.438 | 0.651 |
| 16 | 0.514 | 0.708 | 0.423 | 0.626 |
| 17 | 0.497 | 0.683 | 0.410 | 0.605 |
| 18 | 0.482 | 0.660 | 0.398 | 0.586 |
| 19 | 0.468 | 0.640 | 0.387 | 0.568 |
| 20 | 0.455 | 0.621 | 0.376 | 0.552 |
| 21 | 0.443 | 0.604 | 0.367 | 0.537 |
| 22 | 0.432 | 0.588 | 0.358 | 0.523 |
| 23 | 0.422 | 0.573 | 0.350 | 0.510 |
| 24 | 0.413 | 0.559 | 0.342 | 0.498 |
| 25 | 0.404 | 0.547 | 0.335 | 0.487 |
| 26 | 0.396 | 0.535 | 0.328 | 0.477 |
| 27 | 0.388 | 0.524 | 0.322 | 0.467 |
| 28 | 0.380 | 0.513 | 0.316 | 0.458 |
| 29 | 0.373 | 0.503 | 0.310 | 0.449 |
| 30 | 0.367 | 0.494 | 0.305 | 0.441 |
| 40 | 0.316 | 0.422 | 0.263 | 0.378 |
| 50 | 0.281 | 0.375 | 0.235 | 0.337 |
| 60 | 0.256 | 0.341 | 0.214 | 0.306 |
| 70 | 0.237 | 0.314 | 0.198 | 0.283 |
| 80 | 0.221 | 0.293 | 0.185 | 0.264 |
| 90 | 0.208 | 0.276 | 0.174 | 0.248 |
| 100 | 0.197 | 0.261 | 0.165 | 0.225 |
| 200 | 0.139 | 0.183 | 0.117 | 0.165 |
| 500 | 0.088 | 0.116 | 0.074 | 0.104 |
| ∞ | 0 | 0 | 0 | 0 |

例：下表中的数据是为了确定在内燃机中使用不同的金属轴瓦时，转轴的平均磨损率是否不同而收集的。

表 2　在给定时间后的转轴磨损　　　　单位：0.001毫米

| 转 轴 | 磨 损 | | 差 |
|---|---|---|---|
| i | 铜一铅 $X_i$ | 白色金属 $Y_i$ | $d_i = X_i - Y_i$ |
| 1 | 88.90 | 38.10 | 50.80 |
| 2 | 50.80 | 33.02 | 17.78 |
| 3 | 119.38 | 114.30 | 5.08 |
| 4 | 71.12 | 63.50 | 7.62 |
| 5 | 165.10 | 114.30 | 50.80 |
| 6 | 55.88 | 43.18 | 12.70 |
| 7 | 63.50 | 45.72 | 17.78 |
| 8 | 147.32 | 83.82 | 63.50 |
| 9 | 106.68 | 58.42 | 48.26 |
| 总 数 | 868.68 | 594.36 | 274.32 |

技术特性

| 统计项目 | 计算 |
|---|---|
| 样本大小：<br>$n = 9$ | |
| 观测值的和：<br>$\sum_i X_i = 868.68$<br>$\sum_i Y_i = 594.36$ | $\bar{d} = \frac{1}{9}(868.68 - 594.36) = 30.48$ |
| 差的和：<br>$\sum_i d_i = 274.32$ | $S_d^2 = \frac{1}{8}\left[12\,399.9752 - \frac{(274.32)^2}{9}\right]$<br>$= 504.837\,7$<br>$\hat{\sigma}_d = \sqrt{504.837\,7} = 22.468\,6$ |
| 差的平方和：<br>$\sum_i d_i^2 = 12\,399.975\,2$ | $\frac{t_{0.995}}{\sqrt{9}} = 1.118$ |
| 给定值：<br>$d_0 = 0$<br>自由度：<br>$\nu = 8$<br>显著性水平：<br>$\alpha = 0.01$ | $A_2 = 1.118 \times 22.4686$<br>$= 25.119\,8$<br>$\doteq 25.12$ |

结果：

总体均值 $D$ 与给定值零的比较

双侧情形：

$|\bar{d} - d_0| = 30.48 \quad 25.12$

在显著性水平 1 %下，拒绝两种金属轴瓦磨损率相等的这个假设

## 4 第Ⅱ类错误

当原假设正确时，拒绝此假设的概率至多等于显著性水平 $\alpha$。当原假设正确时，拒绝此假设称为犯第Ⅰ类错误。因此，$\alpha$限定了犯这类错误的风险。

另一方面，可能犯第Ⅱ类错误，即原假设不正确时接受此假设。当原假设不正确时，拒绝它的概率 $1-\beta$称为检验的功效。因此，犯第Ⅱ类错误的概率为 $\beta$。

当已知样本大小 n 和犯第Ⅰ类错误的概率时，犯第Ⅱ类错误的概率不仅依赖于差$d_i = X_i - Y_i$的总体均值$D$（对于$D$，可假定不同的备择假设），而且依赖于这些差的标准差$\sigma_d$。此标准差一般是未知的，当 n 小时，样本仅提供一个粗劣的估计。其结果使得，确定犯第Ⅱ类错误的概率的上限是不可能的。

下列各图，在假设$H_0$（$D \leqslant 0$）的单侧检验的情形，对于不同的 n 值以及显著性水平0.05和0.01，分别给出了检验功效 $1-\beta$和总体均值除以标准差 $D/\sigma_d$之间的关系。

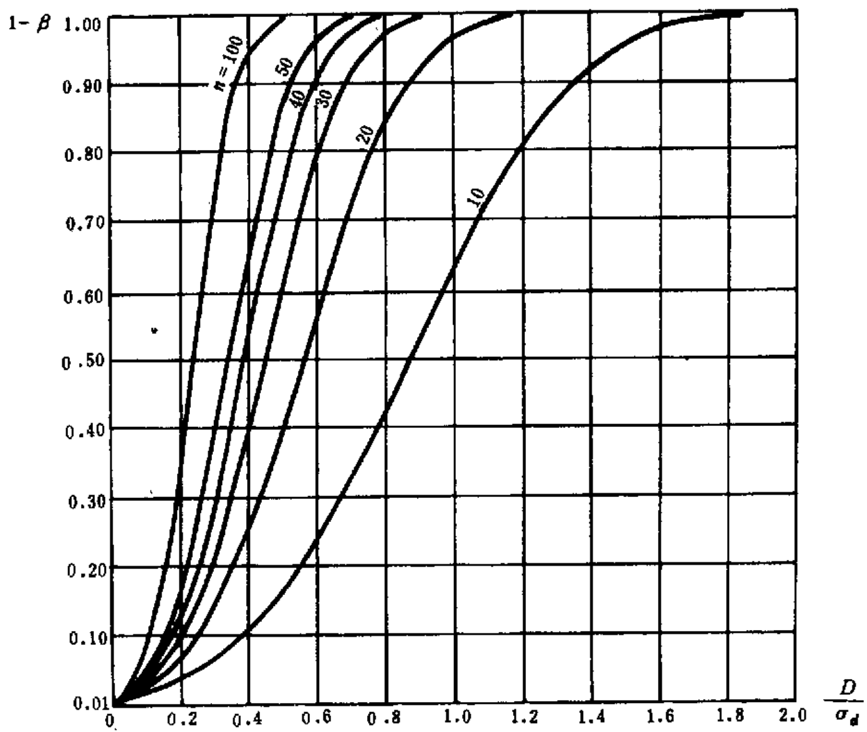


图1　单样本 $t$ 检验的功效（单侧）$\alpha = 0.01$

从图 1 和图 2 中可得出以下结论：

a.　检验功效是由差的总体均值与标准差之比、显著性水平 $\alpha$ 和样本大小 n 唯一确定的。

b.　功效函数是差的总体均值的严格（递）增函数。

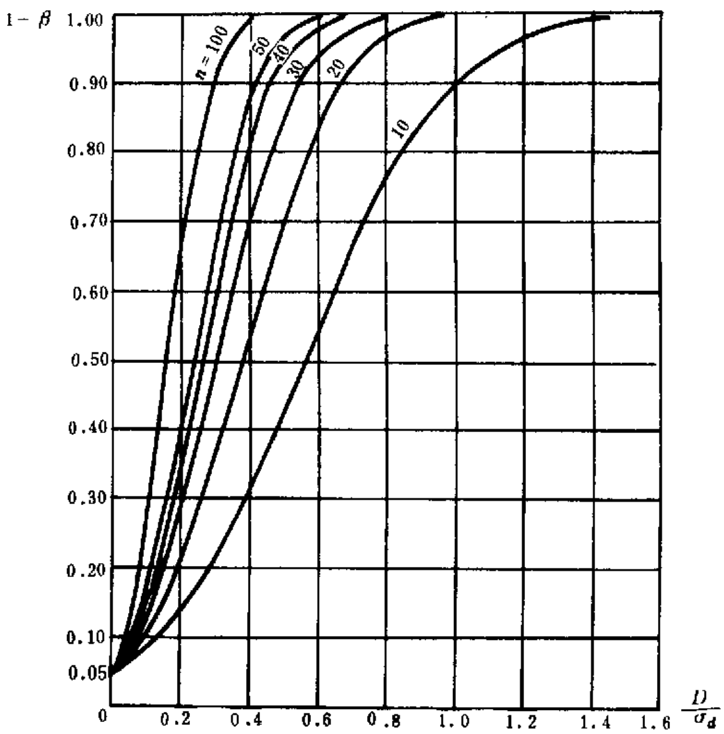


图 2　单样本 $t$ 检验的功效（单侧）$\alpha=0.05$

若$D>0$，$\alpha \neq 0$和1，则功效函数仍是样本大小n和显著性水平$\alpha$的严格（递）增函数。

c.　对于显著性水平0.05和样本大小n＝50，当差的真正均值超过差的标准差的一半(即$\frac{D}{\sigma_d}>0.5$)时，功效至少为0.95。当 n＝20和$D/\sigma_d \geqslant 0.78$时，功效也至少为0.95。

**附加说明：**

本标准由电子工业部标准化研究所提出。

本标准由电子工业部标准化研究所、科学院系统科学研究所和哈尔滨工业大学共同起草。