浙 江 林 学 院 学 报 2005, **22**(1): 119～122
Journal of Zhejiang Forestry College

文章编号: 1000-5692(2005)01-0119-04

均值差异性比较试验次数的确定方法

黄必恒, 管 宇, 吴志松
(浙江林学院 理学院, 浙江 临安 311300)

摘要: 确定适宜的试验重复次数一直是野外试验关注的问题。利用计算机模拟试验, 从控制犯第二类错误的概率来研究两总体均值比较试验时的重复次数问题, 求得在已知历史资料情况下样本容量的计算公式。建议初次试验组内重复次数不宜少于10次。表2参13

关键词: 均值差异性比较; 试验设计; 重复次数
中图分类号: S11^{+}4; Q332 文献标识码: A

试验设计的3个基本原则之一是试验要有重复。重复试验的目的主要是为了估计试验误差和提高试验精度。试验误差是客观存在的, 同一处理在相同条件下进行重复试验, 所得结果不尽相同, 它们之间的差异就是试验误差, 它是由试验中的随机因素所引起的。在对试验结果进行统计分析时, 要利用试验误差来检验各水平之间是否存在着显著差异。根据样本均值的标准差公式 $s_{\bar{x}} = s/\sqrt{n}$, 可知, 试验的重复次数越多, 样本均值的标准差 $s_{\bar{x}}$ 越小, 试验分析的结果越可靠。试验不可避免地会受到外界环境的干扰, 特别是野外试验。周围环境不可能保持一致, 单一试验容易受异常现象的干扰, 使得试验结果不精确。设计了多次重复后可以削弱外界环境的干扰, 因为随机现象的本质会在大量的重复试验中显示出来。单就试验精确性来说, 重复次数越多越好。但是, 试验次数多了会增加试验的成本, 扩大试验地的范围, 这在客观条件上是不允许的。在实际试验中由于受客观条件的限制, 重复次数往往偏少, 使得试验结果不可靠。那么试验的重复次数究竟以多少为宜呢?这是一个比较复杂的问题, 不同的试验设计有不同的要求。文献[1]和文献[2]中有较详细的介绍。本文用计算机模拟试验来探讨重复试验的次数。

1 根据统计原理确定适宜的重复次数

在假设检验中易犯2类错误, 一是原假设 H_0 是正确的, 经检验后被否定了, 称为犯第一类错误, 犯第一类错误的概率为 α, 亦即假设检验的显著性水平 α; 二是原假设 H_0 是不正确的, 经检验后被接受了, 称为犯第二类错误, 犯第二类错误的概率记为 β。在假设检验中总是根据实际问题的要求, 选取适当的显著性水平 α, 以控制犯第一类错误的概率。而犯第二类错误的概率 β 难以人为控制, 它不仅与 α 有关, 而且与样本容量 n(即重复试验次数)的大小有关。

假设 H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 > \mu_2$, 从两总体中分别抽取容量为 n($n_1 = n_2 = n$)的样本, 那有:

$$P\left\{\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2 + s_2^2)/n}} < t_\alpha\right\} = 1 - \alpha。 \tag{1}$$

收稿日期: 2004-05-26; 修回日期: 2004-08-25
作者简介: 黄必恒, 副教授, 从事概率统计研究。E-mail: bh-huang@126.com

当 H_0 为真时,经检验后接受 H_0 的概率为 $1-\alpha$。若 H_0 不真,即两总体均值 $\mu_1 > \mu_2$ 时有[2]:

$$P\left\{\frac{\bar{x}_1-\bar{x}_2}{\sqrt{(s_1^2+s_2^2)/n}} < t_\alpha\right\} = P\left\{\frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sqrt{(s_1^2+s_2^2)/n}} < t_\alpha - \frac{\mu_1-\mu_2}{\sqrt{(s_1^2+s_2^2)/n}}\right\} = 1-\alpha。$$

这表明当 H_0 不真时接受 H_0 的概率为 $1-\alpha$, 也就是犯第二类错误的概率 $\beta = 1-\alpha$。因此要控制犯第二类错误的概率 β, 只要取定 β 的值, 当要区分两总体均值差 $\mu_1-\mu_2 > \delta$ 时有:

$$t_\alpha - \frac{\delta}{\sqrt{(s_1^2+s_2^2)/n}} \leqslant -t_\beta, \tag{2}$$

所以
$$\sqrt{n} \geqslant \frac{(t_\alpha+t_\beta)\sqrt{s_1^2+s_2^2}}{\delta}。 \tag{3}$$

这里的 δ 一般是根据需要和科研要求由试验者确定, β 值只取单侧。在实际应用中通常取 $\beta = 0.1$ 或 0.2[3]。因而当 α 和 β 确定后, 若根据以往的资料知道 s_1^2 和 s_2^2, 要区分两总体均值差 $\mu_1-\mu_2 > \delta$, 可按上式求得样本容量。但由于预先不知道样本数 n, 无从知道 t 分布的自由度, 式中的 t_α 和 t_β 无法查表得到, 故要用数次试算逐步逼近, 当前后 2 个 n 值接近时就是所求的样本容量。

2　计算机模拟试验

设第 i 个总体第 j 个样本观测值为:

$$x_{ij} = \mu_i + \varepsilon_{ij}\ (i = 1, 2, j = 1, 2\cdots n)。 \tag{4}$$

式中 ε_{ij} 表示随机误差, 它服从正态分布 $N(0, \sigma^2)$, σ^2 是总体方差。假设 $\mu_i\ (i = 1, 2)$ 及 σ^2 已知, 由计算机产生 2 个随机数 D_1 和 D_2[4]。令:

$$x_{ij} = \mu_i + \sigma\sqrt{-2\ln(D_1)}\cos(2\pi D_2)。 \tag{5}$$

得到第 i 个总体第 j 次模拟试验的样本值, 重复 n 次得到 2 个总体的 n 个样本模拟值, 对这 2 组样本模拟值作显著性 t 检验。模拟 k 次, 得到 k 个 t 检验值, 其中小于临界值 t_α 的频率近似等于犯第二类错误的概率 β。例如, 已经取得样本资料 $n_1 = 7$, $n_2 = 8$, $\bar{x}_1 = 0.24$, $\bar{x}_2 = 0.13$, $s_1^2 = 0.0091$, $s_2^2 = 0.0039$, 检验值 $t = 2.68 > t_{0.025}(13) = 2.16$[5]。以这些数据近似代替 μ_1, μ_2, σ_1^2, σ_2^2 用计算机模拟 50 次(取 $n = 8$), 得 t 值:

2.76	6.42	0.43	2.23	5.04	3.04	1.95	1.67	3.61	2.64	0.70	3.81	2.57	2.57
1.52	2.67	1.92	2.59	2.74	4.62	2.22	3.09	4.85	2.60	4.00	2.79	3.70	2.92
0.92	3.36	1.17	4.03	2.86	2.37	4.11	1.40	2.73	4.13	1.86	1.61	5.89	3.74
2.78	1.80	5.02	1.26	2.33	1.44	1.92	2.93						

其中小于临界值 $t_{0.025}(14) = 2.145$ 的有 15 个, 占 30%, t 值的变异为 0.455 1。这说明当 $\alpha = 0.025$ 时, 要区分 $\mu_1 - \mu_2 > 0.24 - 0.13 = 0.11$, 取容量 $n = 8$ 时, 犯第二类错误的概率为 0.30。现在不同容量 $n = 8, 9\cdots16$ 时, 各模拟 1 000 次, 计算检验 t 小于临界值 $t_{0.05}$ 的频数和频率如表 1。从表 1 中可以看出, 随着样本容量的增大, 小于临界值的 t 的个数越来越少, 即犯第二类错误的概率越来越小, 当 $n = 13$ 时, 犯第二类错误的概率已小于 0.10。因此取样本容量为 13, 这时 t 的变异为 0.336 7, t 值相对稳定。若按式(3), 当 $n = 8$, $\alpha = 0.025$, $\beta = 0.10$, $t_\alpha(14) = 2.145$, $t_\beta(14) = 1.345$ 时, 计算得 $n \geqslant 13$。再取 $n = 13$, $t_\alpha(24) = 2.064$, $t_\beta(24) = 1.318$ 代入式(3) 计算得 $n \geqslant 12.37$。这与计算机模拟结果相一致。

3　建立样本容量的回归方程

表 2 是摘自不同概率统计教材的样本资料, 利用计算机模拟试验, 求得当 α 和 β 取不同值时的样本容量。

令
$$v = \frac{|\mu_1 - \mu_2|}{\sqrt{(s_1^2+s_2^2)/2}} = \frac{\delta}{\sqrt{(s_1^2+s_2^2)/2}}, \tag{6}$$

表 1　第二类错误概率与样本容量的关系

Table 1　The relation between the second error probability and sample size

项目	样本容量								
	8	9	10	11	12	13	14	15	16
频数	274	221	172	136	110	82	67	56	37
频率	0.274	0.221	0.172	0.136	0.110	0.082	0.067	0.056	0.037

表 2　样本容量的比较

Table 2　The comparision of different sample size

编号	$\bar{x}_1$	$\bar{x}_2$	s_1^2	s_2^2	$v=\frac{\lvert \bar{x}_1-\bar{x}_2\rvert}{\sqrt{(s_1^2+s_2^2)/2}}$	$\alpha=0.025$ $\beta=0.100$			$\alpha=0.025$ $\beta=0.200$			$\alpha=0.05$ $\beta=0.100$		
						n_1	$\hat{n}_2$	n_1^*	n_2	$\hat{n}_2$	n_1^*	n_2	$\hat{n}_2$	n_2^*
1	0.240	0.130	0.009	0.004	1.364 4	13	13	12	10	10	10	10	10	10
2	79.450	76.230	2.225	3.325	1.933 1	7	7	7	6	6	6	5	5	5
3	998	820	2 653	11 784	2.095 1	6	6	6	5	5	5	5	5	5
4	21.500	18.000	7.505	2.594	1.557 6	10	10	10	8	8	8	8	8	8
5	240	214	548	855	1.002 2	23	22	22	18	17	17	19	18	18
6	0.273	0.138	0.028	0.005	1.035 8	21	21	21	15	16	16	17	17	16
7	14.050	13.850	0.075	0.071	0.740 2	40	39	39	30	30	30	34	33	32
8	86.000	85.700	0.040	0.091	1.170 3	17	17	16	13	13	13	14	13	13
9	880	856	555	1 930	0.680 9	45	46	46	35	36	35	37	38	38
10	176	233	541	2 300	1.512 4	11	11	10	8	9	8	8	8	8
11	20.400	19.400	0.886	0.829	1.080 1	20	19	19	15	15	15	16	16	16
12	238.7	193.6	655.0	1 001.0	1.567 8	10	10	10	8	8	8	8	8	8

建立 n 与 v 之间的回归方程：

$$\begin{cases} \hat{n}_1 = 1.68 + \dfrac{20.62}{v^2}(\alpha = 0.025, \beta = 0.100) \\ \hat{n}_2 = 1.58 + \dfrac{15.84}{v^2}(\alpha = 0.025, \beta = 0.200) \\ \hat{n}_3 = 0.82 + \dfrac{17.44}{v^2}(\alpha = 0.050, \beta = 0.100)。 \end{cases} \tag{7}$$

样本相关系数分别为 $r_1 = 0.998\,7$, $r_2 = 0.999\,0$, $r_3 = 0.997\,7$, 曲线回归关系十分显著。由此回归方程可算得相应的样本容量 $\hat{n}_1$, $\hat{n}_2$, $\hat{n}_3$, 并根据式(3) 计算样本容量记为 n^*, 表中的 n, $\hat{n}$, n^* 十分接近。

一般来说，比值 v 小，也就是试验结果数据变动大，试验的重复次数要多。对于野外试验，由于随机因素的干扰大，数据的变动都比较大，所以野外试验的重复次数多一些为好。

4　几点意见

一般认为适宜的试验重复次数应保证误差项的自由度或 t 分布的自由度至少为 12，否则 F 检验或 t 检验的临界值上升很快，难以呈现显著性[6,7]。实践中重复数的确定主要是凭经验，现在林木田间试验中，组内重复次数一般采用 3～10 次[8～13]。在对试验结果进行统计分析时，只对犯第一类错误的概率 α 加以控制，而犯第二类错误的概率未加控制，亦即取 $\beta = 0.5$, $t_\beta = 0$。这时式(3) 成为 $\sqrt{n} \geqslant \dfrac{t_\alpha \sqrt{s_1^2 + s_2^2}}{\delta}$, 因而试验的重复次数往往偏少，试验分析的结果不是很可靠。

在试验中犯第二类错误的概率也是不能忽视的。为了控制犯第二类错误的概率，如果由历史资料知道两总体的方差，要辨别两总体均值差 $\mu_1 - \mu_2 > \delta$, 可以选取定 α 和 β 的值，由式(3) 或式(7) 求得样本容量。

许多试验是在完全没有任何资料的情况下进行的，我们只能凭经验先取一样本容量进行试验，取得一些样本资料后，再根据这些样本资料按式（3）或（7）求得所需的样本容量，以确定重复次数是否适宜，如果重复次数偏少要补做试验。当然，如果条件许可，初次试验样本容量宜大一些，免得重做试验带来麻烦。笔者建议，初次试验时组内重复次数不宜少于 10 次。

参考文献：

[1] 杨纪珂. 应用生物统计[M]. 北京：科学出版社，1983. 89－98.

[2] 盛聚. 概率论与数理统计[M]. 北京：高等教育出版社，1989. 207－214.

[3] 董时富. 生物统计学[M]. 北京：科学出版社，2002. 82－91.

[4] 杨启帆. 数学建模[M]. 杭州：浙江大学出版社，1999. 41.

[5] 沈恒范. 概率论与数理统计[M]. 北京：高等教育出版社，2003. 205.

[6] 南京林学院树木育种研究室. 树木良种选育方法[M]. 北京：中国林业出版社，1984. 300－393.

[7] 马育华. 试验设计[M]. 北京：农业出版社，1982. 20－22.

[8] 李新国，续九如，朱之悌. 林木田间试验适宜重复数和小区株数的研究[J]. 北京林业大学学报，1993, **15**（4）：103－110.

[9] 陈进卿，龚得明，周以飞，等. 品种多点试验中最优地点数和重复次数的确定[J]. 福建农业大学学报，1994, **23**（3）：266－270.

[10] 陈进卿. 品种试验中最优参试品种个数和重复次数的确定[J]. 福建农业大学学报，1995, **24**（3）：363－366.

[11] 童继平，吴跃进，韩立德，等. 作物品种（系）区域试验单重复、多重复试验效果比较研究[J]. 种子，2002,（5）：18－19.

[12] 孙尚拱. 均匀设计中有重复试验的统计分析[J]. 数理统计与管理，2000,（3）：24－29.

[13] 管宇，黄必恒，吴志松，等. 关于正交试验设计的重复试验问题[J]. 浙江林学院学报，2003, **20**（4）：413－418.

Computation of repeat number of tests comparing mean difference

HUANG Bi-heng, GUAN Yu, WU Zhi-song

(School of Sciences, Zhejiang Forestry College, Lin'an 311300, Zhejiang, China)

Abstract: Determining the optimum repeat number is always a great concern of field test. The computer stochastic simulation test was used to study repeat number for controlling the probability of Type Ⅱ errors and establish a formula calculating sample size in the case of known historical data. And repeat number in one section in first test had better more than 10. [Ch, 2 tab. 13 ref.]

Key words: comparison of mean difference; experiment design; repeat number