

文章编号: 1001-411X(2001)01-0009-04

应用水平法与 t 检验评选火炬松半同胞家系

何昭珩, 钟伟华

(华南农业大学林学院, 广东 广州 510642)

摘要: 探讨了用水平法与简单 t 检验在缺区、缺株、方差不齐性条件下评选出优良家系数目上的差别, 及简单 t 检验的精度问题. 提出了用水平法与简单 t 检验相结合进行家系评选. 从 32 个家系中选出优良家系 11 个, 材积现实增益 35.8%.

关键词: 水平法; t 检验; 火炬松; 半同胞家系

中图分类号: S791.255

文献标识码: A

80 年代初, 针对林木子代测定中参试家系间统计株数不等、缺区、方差不齐性等问题, Hatcher 提出采用水平表现 (简称水平法 Performance Level Analysis) 进行家系评选的统计方法^[1]. 随后传入我国^[2], 并开始应用^[3]. 水平法评选家系有它的优越性, 但也存在如何确定家系入选界线, 入选家系与对照相比是否达到统计学上的显著差异, 精度如何? 围绕这些问题, 我们作了初步研究.

1 材料与方法

1.1 材料

供研究分析的材料是 32 个火炬松半同胞家系 (含 CK), 苗木 17 株以上为第 1 组, 17 株以下的为第 2 组, 1983 年栽植, 8 年生 (含苗龄 7 个月) 材积数据. 试验地设在英德市桥头火炬松种子园, 地势平坦, 土壤条件比较均匀一致, 因育苗时遇上几十年一遇的多雨春季, 导致各家系保苗数多寡悬殊. 因此, 只好按各家系存苗数分成 5 个重复及 2 个重复的随机区组设计, 6 株行式小区, 并把它们布置在同一块试验地里.

1.2 方法

方法 1, 采用水平法与简单 t 检验确定家系入选界线; 方法 2, 完全随机设计的单因素方差分析后进行 t 检验^[4]; 方法 3, 双因素 (单因素随机区组) 方差分析后进行 t 检验.

2 计算公式

2.1 单、双因素方差分析及其 t 检验均采用常规方法^[4~6]

水平法公式如下^[1]:

$$L_i = [(M_i + 2SD - M_P) / 4SD] \times 100\%, MF_i = (4MS \times LM_i) / 100 + MT - 2MS$$

当 $M_i \geq M_P + 2SD$ 时, 令 $L_i = 100$; $M_i \leq M_P - 2SD$ 时, 令 $L_i = 0$. 式中 L_i 、 M_i 、 M_P 、 SD 分别为 i 区组的 i 家系性状得分、平均值、亚群体的平均值和标准差. MF_i 、 LM_i 、 MT 、 MS 分别为 i 家系全部区组的性状平均值、平均得分、群体的平均值和标准差.

2.2 简单 t 检验^[5, 6]

$$T_1 = (M_i - CK) / SE_i = (M_i - CK) / (S_i / \sqrt{f_i}) \quad (1)$$

$$T_2 = (M_i - CK) / \sqrt{2S_i^2 / f_i} = T_1 / \sqrt{2} \quad (2)$$

式中 T_1 、 T_2 均属学生氏 t 分布; (1) 式的 CK 为常量, (2) 式的 CK 为非常量; f_i 为 i 家系株数, (1)、(2) 式的自由度均取 $f_i - 1$; M_i 、 S_i 、 SE_i 分别为 i 家系的平均值、标准差与标准误, CK 是比较标准, 在本试验中 CK 是某一家系的平均值或总平均, 当 CK 为总平均时, 因它远比 M_i 稳定, 故视为常量, 在常规的 t 检验中 S_i 是用各参试家系的公共标准差来代替, 通常自由度用误差项自由度, 但当 CK 为非常量时, 用常规计算的 t 值通常是界于 (1)、(2) 式的 t 值之间, 当无严重缺株时自由度对 t 的临界值的影响很少, 可以忽略, 当有严重缺株时按 (2) 式计算偏于保守.

简单 t 检验是把 (1)、(2) 式计算的结果结合起来判断家系的优劣, 当按 (1) 与 (2) 式检验 M_i 均显著大于 CK 时, 则认为 M_i 肯定大于 CK. 若 CK 非常量, (1) 式检验显著而 (2) 式检验不显著, 则视实际情况而定. 这种方法特别适用于参试家系方差不齐的情况, 方差不齐的常规计算式见文献^[6].

收稿日期: 2000-07-02

作者简介: 何昭珩 (1926-), 男, 副教授.

基金项目: “七五” 国家重点科技攻关 (75-07-01-38); 广东省科委重点课题; 联合国发展计划署项目 (CPR/91/153)

©1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

在方法 2 与方法 3 中按下式计算 t 值^{4,9}：株数, r 是区组数.

$$t=(M_i-CK)/S_e \tag{3}$$

此式即常规的 t 检验(L.S.D 法), 式中, S_e 是标准误, 当 CK 为常量时, 在方法 2 中 $S_e=\sqrt{V_e/f_i}$, 在方法 3 中 $S_e=\sqrt{V_e/r}$, CK 非常量时, 在方法 2、方法 3 中 S_e 分别为 $\sqrt{V_e(1/f_i+1/f_j)}$ 、 $\sqrt{2V_e/r}$, V_e 是方差分析中的误差项均方, f_i 与 f_j 分别是 i 家系与对照的

3 结果与分析

3.1 水平法与简单 t 检验的主要结果

经卡方(χ^2)检验, χ^2 值为 50.15> $\chi^2_{0.05}=44.99$, 方差不齐, 因此作水平法与简单 t 检验分析, 结果列入表 1.

表 1 各家系材积的平均值及其分数与增益¹⁾

Tab. 1 The mean growth, performance level score and real gain of volume for every family

序号 No.	家系 family	株数 tree amount	平均数 mean volume/ tree(m ³)	标准差 Sdev.	与总平均比 compared to overall mean	与 8 号比 compared to family 8		得分 score	增益 gain/ %	
					(T_1)	T_1	T_2		DG_1	DG_2
1	27	12	0.033 39	0.010 56	5.11 *	6.64 *	4.70 *	83.9	60.0	104.3
2	29	6	0.030 62	0.006 59	4.76 *	6.49 *	4.59 *	83.5	59.2	103.3
3	14	12	0.026 07	0.010 26	2.79 *	4.36 *	3.08 *	67.9	31.6	68.2
4	20	11	0.025 83	0.009 75	2.73 *	4.31 *	3.05 *	67.5	30.8	67.1
5	24	12	0.025 97	0.010 02	2.82 *	4.43 *	3.13 *	66.9	29.9	65.9
6	34	6	0.024 34	0.013 07	1.12	2.10 *	1.48	65.0	26.5	61.6
7	32	10	0.025 12	0.007 99	2.89 *	4.74 *	3.35 *	64.7	26.0	60.9
8	26	6	0.024 68	0.011 24	1.50	2.51 *	1.78 *	62.7	22.4	56.3
9	30	5	0.022 84	0.012 30	0.91	1.76	1.25	60.6	18.7	51.6
10	25	8	0.021 48	0.008 82	1.18	2.67 *	1.89 *	56.9	12.2	43.3
11	22	10	0.021 89	0.005 67	2.28 *	4.88 *	3.45 *	56.3	11.2	42.0
12	15	12	0.021 68	0.008 46	1.58	3.49 *	2.47 *	55.7	10.0	40.6
13	18	8	0.021 13	0.006 90	1.36	3.27 *	2.31 *	55.2	9.1	39.4
14	23	12	0.021 15	0.007 30	1.59	3.8 *	2.69 *	54.1	7.2	36.9
15	5	22	0.017 24	0.006 21	-0.34	3.09 *	2.19 *	52.7	4.8	33.9
16	17	10	0.020 36	0.007 34	1.10	3.11 *	2.20 *	52.5	4.4	33.4
17	7	27	0.016 28	0.006 85	-1.16	2.37 *	1.68 *	48.4	-2.8	24.2
18	19	11	0.018 54	0.006 64	0.36	2.69 *	1.95 *	47.9	-3.7	23.0
19	9	25	0.016 46	0.006 39	-1.06	2.59 *	1.83 *	47.8	-4.0	22.7
20	21	6	0.018 53	0.003 49	0.50	3.78 *	2.67 *	47.5	-4.4	22.1
21	33	3	0.018 45	0.012 55	0.09	0.73	0.52	47.3	-4.8	21.6
22	3	20	0.014 77	0.008 33	-1.63	0.87	0.62	46.0	-7.2	18.6
23	11	23	0.015 36	0.006 30	-1.86	1.68	1.19	45.9	-7.3	18.5
24	13	26	0.015 20	0.006 65	-2.00	1.57	1.11	45.7	-7.6	18.1
25	4	26	0.015 24	0.005 25	-1.54	1.25	0.88	44.1	-10.4	14.4
26	1	28	0.014 64	0.005 85	-2.87	1.35	0.96	44.0	-10.7	14.1
27	16	6	0.016 99	0.004 51	-0.45	2.09	1.48	43.4	-11.6	12.9
28	2	18	0.014 18	0.004 01	-3.20	0.91	0.64	42.2	-13.8	10.1
29	10	24	0.014 17	0.005 54	-3.22	0.91	0.64	40.2	-17.3	5.60
30	6	26	0.012 76	0.006 46	-3.99	-0.31	-0.22	37.9	-21.4	0.40
31	8	27	0.013 15	0.006 90	-3.51	0.00	0.00	37.7	-21.7	0.00
32	12	17	0.011 54	0.004 95	-5.23	-1.34	-0.95	31.6	-32.6	-13.9

1) DG_1 是各家系与总平均比较的增益, DG_2 是与对照 8 号比较的增益, 总平均材积=0.017 813 m³; “*” 示该家系在 5% 水准显著大于对照 8 号而被选为优良家系

由表 1 可知, 若按水平法大于总平均水平者保留, 则选入 16 个系, 其平均增益为 22.3%; 若大于对照 8 号者入选, 则选入 30 个家系, 其平均增益为 2.9%. 由此可知, 用水平法可对缺株较多与方差不齐的资料作统一分析, 简便、直观性强, 并可确定入选家系的经济效益, 但由于无差异显著性检验, 故对如何挑选优良家系带有一定的盲目性, 在水平法计算结果基础上作简单 t 检验, 则可克服此缺点. 简单 t 检验按下列方法确定候选系是否入选.

当各家系与总平均比较时按(1)式计算 T 值, 若经检验显著高于总平均者入选. 当各家系与参试家系 8 号比较时, 按(1)式与(2)式计算 T 值, 当某家系按 T_2 与 T_1 值检验均显著高于对照, 则肯定该家系显著高于对照, 可入选; 若按 T_1 值检验显著高于对照, 而按 T_2 值检验则否, 就按实际情况而定.

简单 t 检验表明: 各家系与总平均比较有 7 个家系显著高于总平均. 按水平法与 t 检验结合考虑可以认为得分达 56.3 者入选, 选入 11 个家系, 其增益为 35.8%. 与对照 8 号比较时, 用 T_2 检验有 18 个家系的材积显著高于 8 号, 但若用 T_1 检验, 则有 19 个家系的材积显著高于 8 号, 比 T_2 检验结果仅多选入 1 个家系, 通常宁可犯多选入个别家系的错误, 而不愿犯丢失基因资源的错误. 若按水平法平均分数排序, 结合 t 检验, 可以认为家系平均分数至少要达到 47.5 才可入选. 这时, 共有 20 个家系可以入选, 增益为 16.5%, 虽然这 20 个家系中只有 18 个家系肯定显著高于对照 8 号, 有 2 个家系(34 号与 30 号)因

其标准误太大, 平均值不够稳定而没有达到显著水平, 可暂时保留, 这种平均材积大, 方差大的家系虽不宜推广, 但有系内选择的较大可能性, 仍为育种的好材料. 由此可知: 将水平法与简单 t 检验结合起来, 可提供更多对育种有益的信息.

3.2 简单 t 检验精度

取本试验的 13 个家系 5 次重复的全部试验木的材积分别采用方法 1、2、3 三种方法作 t 检验, 这些家系的方差是齐性的, 且无缺区. 其主要结果见表 2.

可见方法 1、2 均选入 4 个家系, 其筛选结果完全相同, 但按方法 3 则仅选入 1 个家系, 其主要原因是方法 3 的家系标准误偏大, 统计推断的精度, 在给了估计的可靠性(P)后, 主要取决于误差项的自由度与标准误(S_e), 但当自由度大于 10 时, 则自由度对精度的影响较小, 这时, 统计推断的精度主要由标准误决定, 若以 C 表示统计推断的精度, 则 C 与标准误 S_e 有下列关系:

$$C=1-t(S_e)/M_i$$

可见, 当 t 与 M_i 固定不变或变化很小时, 则 C 的大小取决于家系的标准误 S_e , 当 S_e 接近于 0 时, 则 C 接近于 1, 即其精度几乎达到百分之百.

由表 2 可知, 方法 3 的标准误明显大于方法 1、方法 2, 故可以断言简单 t 检验的精度至少不会低于以小区平均值为单位的随机区组统计分析结果, 就其原因, 主要是方法 3 的重复数为区组数, 方法 1 的重复数为家系的总株数, 通常家系株数远大于区组数.

表 2 3 种 t 检验法的比较
Tab. 2 Comparison for the result of three methods of t test with family 8 as the control

家系号 family	简单 t 检验 simple t -test		单因方差分析下的 L. S. D. 法 t -test under one factor analysis of variance		双因方差分析下的 L. S. D. 法 t -test under two factors analysis of variance	
	标准误	t 值	标准误	t 值	标准误	t 值
	S. E.	t value	S. E.	t value	S. E.	t value
5	0.143 1	2.78 *	0.141 1	2.82 *	0.182 8	2.00 *
9	0.123 9	2.67 *	0.132 4	2.49 *	0.182 8	1.64
2	0.090 1	2.98 *	0.156 0	1.72 *	0.182 8	1.44
7	0.112 3	2.49 *	0.127 4	2.20 *	0.182 8	1.23
1	0.147 1	0.83	0.125 1	0.98	0.182 8	0.74
3	0.166 2	0.40	0.148 0	0.44	0.182 8	0.82
11	0.158 8	1.02	0.138 0	1.17	0.182 8	0.59
4	0.134 5	1.30	0.129 8	1.35	0.182 8	0.50
13	0.120 2	0.80	0.129 8	0.74	0.182 8	0.55
8(CK)	0.128 8	0.00	0.127 4	0.00	0.182 8	0.00
10	0.139 8	0.80	0.135 1	0.83	0.182 8	-0.02
6	0.125 0	-0.51	0.129 8	-0.50	0.182 8	-0.29
12	0.152 8	-0.61	0.160 6	-0.58	0.182 8	-0.92

4 结论与讨论

4.1 水平法能够对缺株严重或参试家系方差不具齐性的试验作统一分析,它不需剔除数据,补入数据或数据变换,能够充分利用试验提供的信息,客观、简便、直观性强,但若仅用水平法进行家系筛选,由于无统计的显著性检验,而使筛选优良家系的工作带有盲目性。

4.2 简单 t 检验与 L.S.D. 法比较,它优先选入方差小的家系,易于拒绝接受方差大的家系,故若单独使用简单 t 检验评选家系,当各家系的标准误差变化较大时往往会将高产遗传变异大的家系或组合剔除;把水平法与简单 t 检验结合起来可以使其优缺点互补。这时,若以平均水平为据则可选入 11 个家系,其平均增益为 35.8%,若以对照 8 号的水平为据则可选入 20 个家系,其平均增益为 16.5%。

4.3 简单 t 检验的精度至少不低于以小区平均值为单位计算的完全随机区组方差分析后的 L.S.D 法。以小区平均值为单位计算时,其标准误差偏大的主要原因是其重复数偏小,在简单 t 检验中重复数是家系株数,在随机区组方差分析中重复数是区组数,通常是区组数远小于各家系的总株数。本文虽仅就英德 1983 年生子代材积作典型分析,但在林业试验中,缺株与方差不齐是经常遇到的,故其结论具有普遍意义。此法特别适用于小区株数不等、方差不齐的情形。

4.4 简单 t 检验原则上是对 1 对平均值作显著性检验,当同时对多个家系作显著性检验时,可能使统计推断结论的危险率增大,而多选入个别并不显著大

于对照的家系,但这通常对育种工作不会带来什么害处。当要严格要求筛选家系时,只要适当变换显著性水平即可,如显著性水平为 0.001 时,则可保证按简单 t 检验的 50 个家系同时达到显著的危险率低于 5%。

4.5 有时在育种中,种子数不齐,可采用区组数不等,小区株数不等的随机区组设计,尽可能使所有的家系包括在同一区组中,使多些区组含大多数处理。这时,由于在某些区组内各家系株数不相等,适于用方法 1。

4.6 考虑到 t 检验是为了测定某优系是否显著大于对照,故用单侧测验法。若用双侧测验,上述结论也基本是正确的,只不过入选的家系数可能少些。

参考文献:

- [1] HATCHER A R. The use of progeny test data for evaluation of specific crosses and clones[A] . Anon. Tree improvement short course; No 30[C] , Ralugh N C; N C S U press. 1981. 184—187.
- [2] BERNARD O. Statistics in research[M] . Iowa: The Iowa State University Press. 1963. 119—122.
- [3] 广东省杉木子代测定协作组. 杉木种子园子代测定的初步研究[J] . 中南林学院学报, 1989, 9(2): 175—183.
- [4] 莫惠栋. 农业试验统计[M] . 上海: 上海科学技术出版社, 1984. 90—95, 151—206.
- [5] 北京林学院. 数理统计[M] . 北京: 中国林业出版社, 1980. 148—204, 278—281, 289—291.
- [6] 郭祖超. 医用数理统计[M] . 北京: 人民卫生出版社, 1984. 202—205.

Usage of Performance Level Analysis and t Test in Family Selection

HE Zhao-heng, ZHONG Wei-hua

(College of Forestry, South China Agric. Univ., Guangzhou 510642 China)

Abstract: The paper approached the difference of the number of selected superior family between the methods of performance level analysis and simple t -test, under the condition of missing plots, missing trees and variance heterogeneity. The accuracy of simple t -test was analyzed. It was put forward that family selection can be carried out by combining performance level analysis and simple t -test. Eleven superior families were selected from the 32 tested families and the real gain for volume was estimated to be 35.8%.

Key words: performance level analysis; t -test; loblolly pine; half-sib family

【责任编辑 周志红】