

用小区平均值估算单株遗传力的方法

黄少伟 钟伟华

(华南农业大学林学院, 广州 510642)

摘要 子代测定中以小区平均值进行方差分析及估算遗传力, 单株遗传力的结果误差较大, 用小区平均值作方差分析, 外加计算小区内方差及调和平均株数, 进而计算方差成分及遗传力, 其结果与采用单株数据估算的结果相近, 即使在试验数据严重不平衡的情况下, 也不例外. 该法可有效节省计算机资源及时间. 本文以实例详细介绍了有关的计算方法.

关键词 子代测定; 方差分析; 方差成分; 遗传力; 小区平均值

中图分类号 S 722. 89

估算方差成分和遗传力是遗传测定中的重要内容. 单株选择是林木遗传改良的重要内容之一. 要预期单株选择效果, 则需了解单株遗传力, 应用单株观测数据进行估算. 林木遗传育种工作者最常遇见的问题是参试家系间单株数据多寡不一, 对于大规模的、长周期的田间试验来说, 尤其如此. 目前已有多种估算不平衡单株数据方差成分的方法, 如亨德森氏 (Henderson's) 方法 I (HM1 或 TYPE I)、TYPE III 极小方差二次估计 (MIVQUE), 极小标准二次无偏估计 (MINQUE), 最大似然法 (ML) 和限制性最大似然法 (REML) 等 (Corbell et al, 1976; Searle et al, 1992; Huber et al, 1994), 各法所得结果不同. Huber 等 (1994) 用计算机模拟数据对以上 5 种估算方差成分的方法进行比较, 结果认为, 用单株数据估算方差成分时, 以限制性最大似然法 (REML—restricted maximum likelihood) 为佳, 该法适用于不同类型和不同程度的数据缺失, 计算结果稳定性高. 随着计算机应用水平的提高和普及, 上述各种方法在林木育种中的应用日益广泛. SAS/STAT 统计分析软件的一般线性模型 (GLM) 及方差成分估计 (VARCOMP) 程序是不均衡数据方差分析及方差成分估计的有力工具. 然而, 使用 GLM 或 VARCOMP 处理单株数据, 需要占用计算机大量的内存和时间 (SAS Institute Inc, 1990; Stonecypher, 1992), 且由于出现多种平方和而使结果复杂化. 而用小区平均值进行方差分析及估算方差成分, 虽可以克服这些弊端, 但不能估算单株遗传力. Stonecypher (1992) 认为, 较实用的办法是用小区平均值作方差分析, 另外独立计算机误差均方和小区调和平均株数, 可以解决这些问题, 但并未介绍有关的计算公式. 本文根据有关文献, 介绍该法有关计算公式, 并用实例展示运算过程, 进而将估算结果与 3 种基于单株数据的估算方法的结果进行对比, 论证该法的有效性.

1 统计原理与计算方法

子代测定中较常用的试验设计为随机完全区组设计, 用单株数据及小区平均值进行方差分析时, 统计模型分别为:

$$\text{单株数据: } Y_{ijk} = \mu + B_i + F_j + BF_{ij} + \epsilon_{ijk},$$

小区平均: $Y_{ij} = \mu + B_i + F_j + BF_{ij} + \epsilon_{ij} / n$,

其中: Y_{ijk} : 第 i 区组第 j 家系第 k 个单株的观察值(下同); Y_{ij} : 第 i 区组第 j 家系的小区平均值(下同); μ : 总平均值, 固定效应; B_i : 第 i 区组的固定效应值, $i = 1, 2, \dots, b$; F_j : 第 j 家系的随机效应值, $E(F_j) = 0, V(F_j) = \sigma_F^2, j = 1, 2, \dots, f$; BF_{ij} : 第 i 区组第 j 家系的随机互作效应, $E(BF_{ij}) = 0, V(BF_{ij}) = \sigma_{BF}^2$; ϵ_{ijk} : 第 ij 小区第 k 单株的机误, $E(\epsilon_{ijk}) = 0, V(\epsilon_{ijk}) = \sigma_E^2, k = 1, 2, \dots, n$; ϵ_{ij} : 第 ij 小区 n 个单株的机误之和.

多地点试验时, 统计模型分别为:

单株数据: $Y_{ijkl} = \mu + S_i + B_{ij} + F_k + SF_{ik} + BF_{ijk} + \epsilon_{ijkl}$,

单株数据: $Y_{ijk} = \mu + S_i + B_{ij} + F_k + SF_{ijk} + \epsilon_{ijk} / n$,

其中: Y_{ijkl} : 第 i 地点第 j 区组第 k 家系第 l 个单株的观察值(下同); Y_{ijk} : 第 i 地点第 j 区组第 k 家系的小区平均值(下同); μ : 总平均值, 固定效应; S_i : 第 i 地点环境引起的固定效应值, $i = 1, 2, \dots, s$; B_{ij} : 第 i 地点第 j 区组的固定效应值, $j = 1, 2, \dots, b$; F_k : 第 k 家系的随机效应值, $E(F_k) = 0, V(F_k) = \sigma_F^2, k = 1, 2, \dots, f$; SF_{ik} : 第 i 地点环境与第 k 家系的随机互作效应, $E(SF_{ik}) = 0, V(SF_{ik}) = \sigma_{SF}^2$; BF_{ijk} : 第 i 地点内第 j 区组与第 k 家系的随机互作效应, $E(BF_{ijk}) = 0, V(BF_{ijk}) = \sigma_{BF}^2$; ϵ_{ijkl} : 第 ijk 小区组第 l 单株的机误, $E(\epsilon_{ijkl}) = 0, V(\epsilon_{ijkl}) = \sigma_E^2, l = 1, 2, \dots, n$; ϵ_{ijk} : 第 ijk 小区 n 个单株的机误之和.

半同胞子代测定单株数据方差分析, 单株及家系平均狭义遗传力(分别以 h_S^2 和 h_F^2 表示)的计算公式如下(Wright, 1976):

$$h_S^2 = 4\sigma_F^2 / (\sigma_E^2 + \sigma_{BF}^2 + \sigma_{SF}^2 + \sigma_F^2), h_F^2 = \sigma_F^2 / (\sigma_E^2 / sbn + \sigma_{BF}^2 / sb + \sigma_{SF}^2 / s + \sigma_F^2).$$

以上为多地点试验的计算公式, 单地点试验的计算公式为:

$$h_S^2 = 4\sigma_F^2 / (\sigma_E^2 + \sigma_{BF}^2 + \sigma_F^2), h_F^2 = \sigma_F^2 / (\sigma_E^2 / bn + \sigma_{BF}^2 / b + \sigma_F^2).$$

用小区平均值进行方差分析, 小区内机误被忽略, 即机误均方 σ_E^2 被忽略, 单株遗传力被夸大(Wright, 1976), 为此, 必须单独计算 σ_E^2 , 计算公式如下:

$$\text{单地点: } \sigma_E^2 = \frac{b}{i=1} \sum_{j=1}^f \sum_{k=1}^n (Y_{ijk} - Y_{ij})^2 / (N - bf) \text{ 或 } \sigma_E^2 = \frac{b}{i=1} \sum_{j=1}^f \sum_{k=1}^n (Y_{ijk} - Y_{ij} / n_{ij})^2 / (N - bf).$$

其中: Y_{ij} : 第 ij 小区各单株观察值之和, n_{ij} : 第 ij 小区株数, N : 试验全林总株数.

$$\text{多地点: } \sigma_E^2 = \sum_{i=1}^s \sum_{j=1}^b \sum_{k=1}^f \sum_{l=1}^n (Y_{ijkl} - Y_{ijk})^2 / (N - sbf)$$

$$\text{或 } \sigma_E^2 = \sum_{i=1}^s \sum_{j=1}^b \sum_{k=1}^f \sum_{l=1}^n (Y_{ijkl} - Y_{ijk} / n_{ijk})^2 / (N - sbf),$$

其中: Y_{ijk} : 第 ijk 小区各单株观察值之和, n_{ijk} : 第 ijk 小区株数, N : 全部试验林的株数总和. 结果, 相应的期望均方公式如下:

单地点:	期望均方	多地点:	期望均方
变异来源		变异来源	
家系	$\sigma_E^2 / n_h + \sigma_{BF}^2 + b\sigma_F^2$	家系	$\sigma_E^2 / n_h + \sigma_{BF}^2 + b\sigma_{SF}^2 + sb\sigma_F^2$
区组 × 家系	$\sigma_E^2 / n_h + \sigma_{BF}^2$	地点 × 家系	$\sigma_E^2 / n_h + \sigma_{BF}^2 + b\sigma_{SF}^2$
机误	σ_E^2 / n_h	区组 × 家系	$\sigma_E^2 / n_h + \sigma_{BF}^2$
		机误	σ_E^2 / n_h

式中“区组×家系”项,计算机以“机误”形式输出, n_h 为小区调和平均株数,其计算公式为:

$$\text{单地点: } n_h = bf / \sum_{i=1}^b \sum_{j=1}^f (1/n_{ij}), \text{ 多地点: } n_h = sbf / \sum_{i=1}^s \sum_{j=1}^b \sum_{k=1}^f (1/n_{ijk}).$$

在参试家系单株数据均等的情况下, $n_h = n$, 则上述两式期望均方均乘以小区株数 n , 即为平衡单株数据的期望均方。

对不同方法估算的遗传力进行差异显著性检验, 首先计算各方法遗传力的置信区间(北京师范大学生物系、数学系数量遗传研究组, 1978), 若两个遗传力的置信区间相重叠, 说明两个遗传力间不存在显著差异, 否则, 两者间差异显著。遗传力置信区间的上限

$$\bar{\rho} = \frac{[1 + (bn - 1)h^2] - [(1 - h^2)/F_{\alpha/2}(df_2, df_1)]}{[1 + (bn - 1)h^2] + [(bn - 1)(1 - h^2)/F_{\alpha/2}(df_2, df_1)]},$$

下限

$$\rho = \frac{[1 + (bn - 1)h^2] - (1 - h^2)F_{\alpha/2}(df_1, df_2)}{[1 + (bn - 1)h^2] + (bn - 1)(1 - h^2)F_{\alpha/2}(df_1, df_2)},$$

其中: b : 子代测定的区组数, n : 小区单株数, h^2 : 估算所得遗传力; $F_{\alpha/2}(df_1, df_2)$: 以 df_1 为第一自由度, df_2 为第二自由度, 信度为 $1 - \alpha$ 的 F 分位数; $F_{\alpha/2}(df_2, df_1)$: 以 df_2 为第一自由度, df_1 为第二自由度, 信度为 $1 - \alpha$ 的分位数; df_1 : 方差分析中的家系自由度; df_2 : 小区平均值方差分析中的机误自由度或单株数据方差分析中的区组×家系自由度。

2 实例与分析

2.1 试验数据

数据来自广东英德市火炬松种子园一片 9 年生火炬松半同胞子代测定林, 随机完全区组设计, 30 个家系(含 1 个对照), 7 次重复, 6 株行式小区。试验林保存率为 85.4%。为了比较, 数据处理成 3 种不平衡状态, 即(1)保持原状态, 保存率 85.4%, (2)删除 1~5 区组各小区第 2 株数据, 保存率为 74.6%, (3)删除 1-20 号家系各小区第 2、5 株数据, 保存率为 65.4%。

2.2 随机效应差异显著性检验

表 1 3 种保存率下单株数据方差分析结果¹⁾

变异来源	df	85.4% ²⁾						74.6% ²⁾						
		均方差			方差比			均方差						
		树高	胸径	材积	树高	胸径	材积	df	树高	胸径	材积			
家系	29	3.034	16.605	0.74×10 ⁻³	2.23**	2.32**	2.06**	29	2.628	16.833	0.73×10 ⁻³			
区组×家系	17	1.363	7.149	0.36×10 ⁻³	2.46**	1.64**	2.04**	17	1.235	6.850	0.34×10 ⁻³			
机误	86	0.553	4.365	0.18×10 ⁻³				73	0.543	4.348	0.18×10 ⁻³			
变异来源	df	74.6% ²⁾						65.4% ²⁾						
		方差比			df	均方差			方差比					
		树高	胸径	材积		树高	胸径	材积	树高	胸径	材积			
家系		2.13**	2.46**	2.15**	29	2.591	15.944	0.70×10 ⁻³	2.18**	2.42**	2.11**			
区组×家系		2.28**	1.58**	1.85**	17	1.188	6.594	0.33×10 ⁻³	2.13**	1.43**	1.76**			
机误					61	0.559	4.601	0.19×10 ⁻³						

1) * 和 ** 分别表示在 5% 和 1% 水准上显著; 2) 保存率

采用 SAS/STAT 的 GLM 过程进行单株数据方差分析, 取 TYPE III 平方和, 结果(表 1), 家系间树高、胸径和材积均有极显著差异, 区组×家系互作效应也极显著。

2.3 方差成分及遗传力估算

小区平均值方差分析, 附加计算 σ_E^2 及 n_h , 依上述方法可列出相应的期望均方公式, 并估算方差成分. 用 SAS/STAT 的 GLM 过程进行单株数据方差分析, 取 TYPE III 平方和, 据计算机输出的期望均方公式可估算方差成分. 用 VARCOMP 过程进行基于单株数据的方差成分估算, 取 TYPE I 和 REML 两种方法, 计算机输出两种方法的方差成分和 TYPE I 的期望均方. 期望均方及方差成分估算结果分别见表 2、3.

用 Wright(1976)的方法计算树高、胸径及材积 3 性状的单株及家系遗传力(表 4). 由表 4 可见, 小区平均值方差分析, 附加计算 σ_E^2 及 n_h 所得遗传力与单株数据为据估算的结果相近.

表 2 期望均方公式

保存率	85.4%	74.5%	65.4%
	$\sigma_E^2/4.487 + \sigma_{BF}^2 + 7\sigma_F^2$	$\sigma_E^2/3.903 + \sigma_{BF}^2 + 7\sigma_F^2$	$\sigma_E^2/3.394 + \sigma_{BF}^2 + 7\sigma_F^2$
小区平均	$\sigma_E^2/4.487 + \sigma_{BF}^2$	$\sigma_E^2/3.903 + \sigma_{BF}^2$	$\sigma_E^2/3.394 + \sigma_{BF}^2$
	$\sigma_E^2/4.487$	$\sigma_E^2/3.903$	$\sigma_E^2/3.394$
	$\sigma_E^2 + 5.234\sigma_{BF}^2 + 35.791\sigma_F^2$	$\sigma_E^2 + 4.606\sigma_{BF}^2 + 31.261\sigma_F^2$	$\sigma_E^2 + 4.034\sigma_{BF}^2 + 27.347\sigma_F^2$
TYPE I	$\sigma_E^2 + 5.093\sigma_{BF}^2$	$\sigma_E^2 + 4.443\sigma_{BF}^2$	$\sigma_E^2 + 3.890\sigma_{BF}^2$
	σ_E^2	σ_E^2	σ_E^2
	$\sigma_E^2 + 4.946\sigma_{BF}^2 + 34.622\sigma_F^2$	$\sigma_E^2 + 4.300\sigma_{BF}^2 + 30.100\sigma_F^2$	$\sigma_E^2 + 3.734\sigma_{BF}^2 + 26.140\sigma_F^2$
TYPE III	$\sigma_E^2 + 4.093\sigma_{BF}^2$	$\sigma_E^2 + 4.443\sigma_{BF}^2$	$\sigma_E^2 + 3.890\sigma_{BF}^2$
	σ_E^2	σ_E^2	σ_E^2

表 3 4 种方法对 3 种保存率估算的方差成分

性状	估算方法	85.4% ¹⁾			74.5% ¹⁾			65.4% ¹⁾		
		σ_F^2	σ_{BF}^2	σ_E^2	σ_F^2	σ_{BF}^2	σ_E^2	σ_F^2	σ_{BF}^2	σ_E^2
树高	小区平均	0.044 4	0.159 5	0.553 3	0.042 5	0.155 6	0.542 5	0.043 5	0.174 0	0.558 6
	TYPE I	0.054 8	0.158 9	0.553 3	0.054 6	0.155 9	0.542 5	0.062 2	0.161 7	0.558 6
	TYPE III	0.048 9	0.158 9	0.553 3	0.047 0	0.155 9	0.542 5	0.054 6	0.161 7	0.558 6
	REML	0.050 6	0.160 7	0.552 8	0.049 3	0.158 3	0.541 7	0.055 5	0.161 2	0.559 6
胸径	小区平均	0.266 8	0.436 7	4.364 5	0.314 9	0.477 3	4.348 0	0.379 8	0.452 8	4.601 1
	TYPE I	0.283 6	0.546 8	4.364 5	0.328 7	0.563 2	4.348 0	0.369 0	0.512 3	4.601 0
	TYPE III	0.275 5	0.546 8	4.364 5	0.334 3	0.563 2	4.348 0	0.360 7	0.512 3	4.601 0
	REML	0.273 6	0.574 6	4.335 6	0.325 1	0.583 8	4.326 3	0.379 5	0.529 1	4.589 6
材积 ($\times 10^{-4}$)	小区平均	0.11	0.32	1.77	0.13	0.30	1.83	0.15	0.36	1.89
	TYPE I	0.12	0.36	1.77	0.14	0.35	1.83	0.15	0.37	1.89
	TYPE III	0.11	0.36	1.77	0.13	0.35	1.83	0.14	0.37	1.89
	REML	0.11	0.37	1.76	0.13	0.36	1.81	0.15	0.37	1.88

1) 保存率

表4 4种方法对3种保存率估算的遗传力

估算方法	85.4% ¹⁾						74.6% ¹⁾						65.4% ¹⁾					
	树高		胸径		材积		树高		胸径		材积		树高		胸径		材积	
	单株	家系	单株	家系	单株	家系	单株	家系	单株	家系	单株	家系	单株	家系	单株	家系	单株	家系
小区平均	0.235	0.524	0.211	0.570	0.199	0.518	0.229	0.502	0.245	0.581	0.223	0.533	0.224	0.474	0.280	0.595	0.246	0.530
TYPE I	0.286	0.586	0.218	0.584	0.207	0.532	0.290	0.575	0.251	0.597	0.235	0.552	0.318	0.584	0.269	0.602	0.248	0.547
TYPE III	0.257	0.559	0.212	0.574	0.200	0.522	0.252	0.538	0.255	0.588	0.226	0.562	0.282	0.551	0.264	0.591	0.239	0.535
REML	0.265	0.555	0.211	0.554	0.200	0.508	0.263	0.537	0.248	0.574	0.227	0.526	0.286	0.544	0.276	0.585	0.251	0.532

1) 保存率

2.4 不同方法所得遗传力的差异显著性检验

计算各遗传力的置信区间进行差异显著性检验, 列于表5. 表5显示, 4种方法估算所得遗传力在5%水准上的置信区间高度重叠, 说明各种方法的估算结果没有显著差异.

表5 4种方法估算的遗传力差异显著性检验¹⁾

估算方法	树高		胸径		材积	
	单株	家系	单株	家系	单株	家系
小区平均	0.136~0.365	0.344~0.631	0.178~0.431	0.465~0.735	0.152~0.391	0.398~0.681
TYPE I	0.210~0.473	0.454~0.726	0.172~0.417	0.473~0.740	0.155~0.392	0.416~0.695
TYPE III	0.181~0.433	0.420~0.699	0.167~0.411	0.461~0.732	0.148~0.382	0.404~0.685
REML	0.183~0.439	0.412~0.693	0.176~0.427	0.455~0.727	0.156~0.397	0.401~0.683

1) 为节省篇幅, 仅列出保存率为65.4%的检验结果, 3种保存率下的检验结果相同. 表中数据为各遗传力在5%水准上的置信区间.

表4.5说明, 用小区平均值方差分析, 附加计算 σ_E^2 及 n_h 的方法估算遗传力, 与单株数据估算的结果有一定差距, 但差距不大, 不存在显著差异, 即使在数据严重不平衡的情况下(保存率仅为65.4%, 缺株程度在家系间差异较大), 其结果与单株数据的估算结果的差距也没有明显增大的迹象, 说明该法是可靠有效的.

3 结论

用小区平均值方差分析, 附加计算机误差均方及小区调和平均株数的方法估算方差成分及遗传力, 其结果与单株数据估算的结果相近, 稳定性不低于单株数据的估算结果, 可有效节约计算机资源及时间. 同时, 在缺株不缺区的情况下, 使方差分析得到唯一的平方和结果.

参 考 文 献

北京师范大学生物系、数学系数量遗传研究组. 1978. 冬小麦数量性状遗传力的初步研究. 遗传学报, 5(2): 145~152

Corbell R R, Searle S R. 1976. A comparison of variance component estimation. Biometrics 32: 779~791

Searle S R, Casella G, McCulloch C E. 1992. Variance Components. New York: John Wiley & Sons Inc. 168~257

Huber D A, White T L, Hodge G R. 1994. Variance component estimation techniques compared for two mating designs with forest genetic architecture through computer simulation. Theor Appl Genet, 88: 236

~ 242

- SAS Institute Inc. 1990. SAS/STAT User's Guide, Version 6; Vol 2. 4th ed. Cary (N. C.); SAS Institute Inc, 891~996, 1661~1673
- Stonecypher R W. 1992. Analyses of forest genetics experiments; Some examples. In: Fins L, Friedman S T, Brotschol J V, eds. Handbook of quantitative forest genetics. Dordrecht; Kluwer Academic Publishers, 195~228
- Wright J W. 1976. Introduction to Forest Genetics. New York; Academic Press Inc, 239~252

USAGE OF ESTIMATION OF SINGLE TREE HERITABILITY WITH PLOT MEANS

Huang Shaowei Zhong Weihua

(Forestry Faculty, South China Agric. Univ., Guangzhou, 510642)

Abstract

For a progeny test, when analysis of variance is undertaken and heritabilities are calculated with plot means, the results of single tree heritability are not desirable. However, if carrying a plot means based analysis of variance with the within-plot variances and the harmonic means of number of individuals per plot to be calculated separately, and then estimating the variance components and heritabilities, the results are similar to those which are estimated with individual observations even when the data set is of serious unbalance. A lot of computer resource and time can be saved under plot means based estimation. The data set from a field experiment was collected to show the detail method and process of this kind of analysis.

Keywords progeny test; analysis of variance; variance component; heritability; plot mean

(上接第 75 页)

TWO NEW GENERA OF ORCHIDACEAE FROM CHINA

Liu Zhongjian¹ Zhang Jinning²

(1 Shenzhen City Wutongshan Nurseries, 518114, 2 College of Forestry, South China Agric. Univ.)

Abstract

Two new genera of Orchidaceae are described; *Luiguishania* Liu et Zhang from Taiwan, China and *Wutongshania* Liu et Zhang from Guangdong, China.

Key words Orchidaceae; *Luiguishania*; *L. taiwanensis*; *Wutongshania*; *W. guangdongensis*