

林木遗传育种中平衡不平衡、 规则不规则试验数据处理技巧

Analysis Skills of Test Data about Balance or Imbalance,
Regular or Irregular in Genetics and Breeding of Forest Tree

齐 明 何贵平 著



中国林业出版社

林木遗传育种中平衡不平衡、 规则不规则试验数据处理技巧

Analysis Skills of Test Data about Balance or Imbalance,
Regular or Irregular in Genetics and Breeding of Forest Tree

齐 明 何贵平 著



中国林业出版社

图书在版编目(CIP)数据

林木遗传育种中平衡不平衡、规则不规则试验数据处理技巧/齐明,何贵平著. —北京:中国林业出版社,2014.9

ISBN 978-7-5038-7633-2

I. ①林… II. ①齐… ②何… III. ①林木育种—遗传育种—研究 IV. ①S722.3

中国版本图书馆CIP数据核字(2014)第205565号

中国林业出版社·生态保护出版中心

责任编辑:李 敏

电 话:(010)83280498

出版 中国林业出版社(100009 北京西城区德胜门内大街刘海胡同7号)

<http://lycb.forestry.gov.cn>

E-mail:lmbj@163.com

发行 中国林业出版社

印刷 北京卡乐富印刷有限公司

版次 2014年10月第1版

印次 2014年10月第1次

开本 787mm×1092mm 1/16

印张 11.75

字数 288千字

定价 36.00元

2009 年, 中国林业出版社出版了齐明同志的专著《林木遗传育种中试验统计法新进展》。在其后的几年里, 同行们给该书提出了一些很好的建议: 该书理论创新, 提出了一个转化理论, 以此为基础, 构建科学的线性模型, 从而克服了负的方差分量, 并使参数估计的可靠性、精确性都有很大的提高, 但它过于复杂, 计算过程繁复, 使用起来不太方便。从而促使我们萌生了研制程序软件的想法。MATLAB 语言具有强大的计算能力和图像处理功能, 我们选择了 Matlab7.0 平台, 来处理平衡不平衡、规则不规则试验资料。经过几年的努力, 林木遗传育种中平衡不平衡、规则不规则试验数据处理程序软件 MLAP 终于研制成功, 并获得国家专利(证书号: 软著登字第 0685204 号)。为了使研究成果能更好地服务社会, 著作者决定出版《林木遗传育种中平衡不平衡、规则不规则试验数据处理技巧》一书, 并用专利来保护该书中的程序不被非法传播。

该书共分五个部分: 第 1 章是转化分析法的统计学基础; 第 2 章是使用本程序软件必备的 M 语言知识; 第 3 章是 MLAP 程序软件的内容和使用; 第 4 章是林木遗传育种中若干试验类型的 M 语言程序; 第 5 章是试验模型选择、数据采集和不平衡数据分析技巧; 书的末尾是附录, 新开发两个常用模型, 示范解决负的方差分量问题。

MLAP 的主要功能: 借助于 Matlab7.0 平台, 对林木遗传育种中获得的平衡不平衡、规则不规则的试验数据进行统计分析, 获得多种遗传参数。分析内容基本涵盖了常规林木遗传育种的各个方面: ①林木子代田间试验; ②林木的选择育种; ③林木的杂交育种和数量遗传; ④林木遗传与育种应用;

⑤品种的稳定性分析等。

MLAP 的主要特点：①统计分析的线性模型是建立在“转化理论”的基础上；②以单株观察值参与统计分析，可以获得更精确更合理的遗传参数；③消灭负的方差分量和遗传相关系数大于 1.0 的现象；④MLAP 具有通俗易懂，内容丰富，功能强大，算法先进，可靠性好，能获得林木遗传育种所需的各种参数；同时这些程序还具有易于扩张和拉伸的功能，适应性广，值得在林木遗传改良领域推广应用。

该书适用于广大林业科研、教学、管理和生产经营者参考使用。

由于作者水平有限，该书中可能有不足之处，欢迎广大读者指正！

中国林业科学研究院亚热带林业研究所

齐明、何贵平

2014 年 4 月

Preface

In 2009, the work of *New Development of Test and Statistics of Genetic Breeding in Forest Trees* of vice professor Qi Ming had been published by China Forestry Publishing House. In subsequent years, Some good advice for the book had been suggested by colleagues: The theory of transformation had been proposed in the work. It is the theoretical innovation, leading to the advent of the work. Different from traditional methods, the work is based on this transformation theory, the construction of the linear model with science, to overcome the negative variance components, and the reliability and accuracy of parameter estimation had been greatly improved. But it is too complicated and cumbersome process, not so easy to use. These reasons have led us to the initiation of development of software ideas. M computer language has a powerful computing power and image processing functions. We choose the Matlab7.0 platform, treatment test data with balanced or unbalanced, irregular or regular. After several years of efforts, program software (MLAP) to treatment test data with balance or unbalanced, regular or irregular has finally succeeded. And application for the national patent (Certificate No. 0685204). To enable research achievements to better serve the community, the item team decided to publish a work. The program in work had been protected by Patent Law to prevent the illegal spread of these programs.

The book is divided into five parts: Chapter 1, statistics foundations of transformation analysis; Chapter 2, prerequisite knowledge of M language when using the MLAP software; Chapter 3, introduction of the program software and use of MLAP program software; Chapter 4, the M language programs of different tests in genetics and breeding of forest trees; Chapter 5, selection on test types and linear models, data collection as well as handling skills about balance or unbalanced, regular or irregular data; The appendix at the end of the book is two models of new development, and it demonstrates how to solve the problem of negative variance components.

Main function of MLAP: by means of platform of Matlab7.0, statistical analysis of test data about balanced or unbalanced, regular or irregular in genetics and breeding of forest trees can be proceeded in order to obtain a variety of genetic parameters. Analysis content had basically covered all aspects of conventional tree genetic breeding: ①Field experiment of progeny in forest tree; ② Selection breeding of tree; ③Cross breeding of tree and quantitative genetics; ④Genetics of tree and its application in breeding; ⑤Stability analysis of forest tree strains and so on.

Main features of MLAP: ①The linear model of statistical analysis is built on the basis of the theory of transformation; ②Individual value is used to statistical analysis, it can gain more precise and more reasonable genetic parameters; ③it can overcome the phenomenon of negative variance components and genetic correlation coefficient greater than 1.0; ④MLAP has a simple, content-rich, powerful, the algorithm advances, good reliability, a variety of parameters required for ge-

netic breeding of forest tree; and also it has features easy to expand and stretch, wide adaptability, to be worth of popularization and application in the field of genetic improvement of forest trees.

This book is suited for the general forestry research, teaching, management and production operators as reference.

Due to the limited level of the authors, there may be deficiencies in the book, Welcome to correct us!

The Research Institute of Subtropical Forestry,
Chinese Academy of Forestry Science: Qi Ming, He Guiping
2014,4

Contents

Preface

Chapter 1 Statistics Foundations of Transformation Analysis /1

- 1.1 Introduction /1
- 1.2 Models and Methods /1
- 1.3 Results and Analysis /3
- 1.4 Conclusions /8

Chapter 2 Prerequisite Knowledge of M Language When Using the MLAP Program Software /9

- 2.1 The Establishment of an Array and Its Various Operations /9
- 2.2 Setting of Matrix /10
- 2.3 The Partition of Matrix and setting of Submatrix /10
- 2.4 Sparse Matrix and Its Relationship with Normal Matrix /11
- 2.5 Operations of the Array /12
- 2.6 Complex Operations of an Array /13
- 2.7 Statistic Analysis of Matrix /13
- 2.8 Relationship between Matrix and Determinant /15
- 2.9 Square Roots of Matrix and Square of Matrix Operations /16
- 2.10 Programming Statements in M Language /16
- 2.11 The Calculation of Symbolic Variable /17
- 2.12 Operations of Symbolic Matrix /19
- 2.13 Production of Matrix with Specific Distributions /21
- 2.14 Accuracy Control and Outputs of Results /21

Chapter 3 Introduction of the Program Software and Use of MLAP Program Software /23

- 3.1 Importance of Research and Development on MLAP Program Software /23
- 3.2 Overall Design of the Software /23
- 3.3 Functions and Features of MLAP /25
- 3.4 Applications of MLAP Software /25

Chapter 4 The M Language Programs in Genetics and Breeding Experiments /31

- 4.1 Analysis Program of Single Factor Randomized Test with Unbalanced Data /31
- 4.2 Analysis Program of Single Factor Randomized Block Design with Unbalanced Data /32
- 4.3 Analysis Program of Two Factors Design with Too Many Repeat Numbers /34
- 4.4 Analysis Program of Unbalanced, Regular or Irregular Nested Design with Three Levels /44
- 4.5 Analysis Program of Factorial Design (Two Factors Randomized Block Design with Unbalanced, Regular or Irregular Data) /52
- 4.6 Analysis Program of Unbalanced, Regular or Irregular Nested Design with Four Levels /62
- 4.7 Analysis Program of Diallel Cross with Unbalanced, Regular or Irregular Data by MINQUE(1) /73
- 4.8 Treatment Technique of Two-level Diallel Cross Tests with Unbalanced, Regular or Irregular Data /78
- 4.9 Analysis Program and Method of Genetic relationship and Selection Index /80
- 4.10 Analysis Program of Breeding Values of Parents and Breeding Values of Individuals /89
- 4.11 Analysis Programme of Strains Stability of Unbalanced Test Data at Many Sites /107
- 4.12 Other Analysis Techniques /118

Chapter 5 Selection on Test Types and Linear Models and Data Collection as well as Treatmental Skills about Unbalanced, Regular or Irregular Data /120

- 5.1 Input formats of Unbalanced Data of ANOVA of Single Factor Randomized Test /120
- 5.2 Input Formats of Unbalanced Data of Single Factor Randomized Block Design /120
- 5.3 Input Formats of Unbalanced Data of Two Factors Design with too Many Repeat Numbers /122
- 5.4 Linear Models and Input Formats of Unbalanced Nested Design of Three Levels with Unbalanced, Regular or Irregular Data /123
- 5.5 Linear Models and Input Formats of Two Factors Randomized Block Design with Unbalanced Data /124
- 5.6 Linear Models and Input Formats of Unbalanced Nested Design with Four Levels /125
- 5.7 Linear Models and Input Formats of Diallel Cross breeding with Unbalanced, Regular or Irregular Data in Forest Trees /128
- 5.8 Input Formats of Analysis of Variance-Covariance for Single Factor Complete Random Block Design with Unbalanced Data /129
- 5.9 Linear Models and Analysis Techniques of Two-level Diallel Cross Tests with Unbalanced, Regular or Irregular Data /130
- 5.10 The Techniques of Unbalanced Data Analysis in Field Test of Forest Trees with Many Sites /132

- 5.11 Analysis Techniques about Unbalanced Data in Genetic Breeding of Forest Trees /132
- 5.12 Analysis Techniques of Genetic Relationship When Test Data were Unbalanced /134
- 5.13 Analysis Techniques for Prediction on Breeding Values by BLP /136
- 5.14 Evaluation Techniques about Breeding Values of Full – sibs Test in Forest Tree by BLUP /137
- 5.15 Evaluation Techniques about High Yield and High Stability When Unbalanced Test Data at Many Sites /138
- 5.16 Analysis Methods and Techniques of the Unbalanced Data from Field Test of Polycross /140

References /142

Appendixes /144

- Appendixes 1 : Analysis Principle of Irregular Nested Design of Three Levels with Unbalanced Data and M Programs /144
- Appendixes 2 : Model Analysis of Reduction Method about Factorial Design with Unbalanced, Regular or Irregular Data and M Programs /152

前 言

第 1 章 转化分析法的统计学基础 /1

1.1 引言 /1

1.2 模型与方法 /1

1.2.1 平衡试验设计及其统计分析模型 /1

1.2.2 平衡试验数据的产生 /2

1.2.3 非平衡试验数据的产生及统计分析 /2

1.2.4 试验数据的转化分析法 /2

1.2.5 Monte Carlo 模拟数据分析结果的评价 /2

1.3 结果与分析 /3

1.3.1 五个试验中，交互项负的方差频率 /3

1.3.2 几个参数效益的比较 /3

1.3.3 转化对遗传力大小的影响 /6

1.4 小结 /8

第 2 章 使用本程序软件必备的 M 语言知识 /9

2.1 数组的建立及其变化操作 /9

2.2 矩阵的合并 /10

2.3 矩阵的剖分与子矩阵的合并 /10

2.4 稀疏矩阵及其与普通矩阵间的联系 /11

2.5 数组的加减乘除运算 /12

2.6 数组的复杂运算 /13

2.7 矩阵的统计分析 /13

2.8 矩阵与行列式间的关系 /15

2.9 矩阵的开方与平方运算 /16

- 2.10 M 语言中的编程语句 /16
- 2.11 符号变量的计算 /17
- 2.12 符号矩阵的运算 /19
- 2.13 特定分布矩阵的产生 /21
- 2.14 精度控制与结果输出 /21

第3章 MLAP 程序软件的内容简介和软件的使用 /23

- 3.1 程序软件 MLAP 研发的意义 /23
- 3.2 软件的总体设计 /23
 - 3.2.1 软件的总体流程图和描述 /23
 - 3.2.2 MLAP 的需求信息 /25
- 3.3 本软件的功能和特点 /25
 - 3.3.1 本软件的功能 /25
 - 3.3.2 本软件的特点 /25
- 3.4 软件的使用 /25

第4章 林木遗传育种中若干试验类型的 M 语言程序 /31

- 4.1 单因素完全随机试验不平衡资料的分析方法 /31
 - 4.1.1 单因素类内观察值不等的方差分析线性模型 /31
 - 4.1.2 不平衡数据处理程序 /31
- 4.2 单因素随机区组试验数据的转化分析法 /32
 - 4.2.1 求因子的离差平方和 /32
 - 4.2.2 求方差分量 /32
 - 4.2.3 对全同胞家系进行 F 检验 /33
- 4.3 两因素有众多重复次数且不平衡的方差分析程序 /34
 - 4.3.1 统计模型 /34
 - 4.3.2 求方差分量 /35
 - 4.3.3 主效因子的显著性 F 检验 /36
- 4.4 三阶巢式设计非平衡试验数据的处理方法 /44
 - 4.4.1 统计模型 /44
 - 4.4.2 方差分析程序 /44
 - 4.4.3 求期望均方系数 /45
 - 4.4.4 求各因子的方差分量程序 /46
 - 4.4.5 各参试因子的 F 检验 /47
- 4.5 平衡不平衡、规则不规则的析因设计数据的方差分析程序 /52
 - 4.5.1 求因子均方 /52
 - 4.5.2 求期望均方系数和方差分量 /54
 - 4.5.3 主要因子的 F 检验 /56
- 4.6 四阶巢式设计平衡不平衡、规则不规则数据的分析程序 /62

- 4.6.1 四阶巢式设计平衡不平衡、规则不规则数据分析的统计模型 /62
- 4.6.2 四阶巢式设计平衡不平衡、规则不规则数据的分析程序 /62
- 4.6.3 自由度及均方求算 /65
- 4.6.4 求因子系数和方差分量 /65
- 4.6.5 因子的显著性 F 检验 /69
- 4.7 采用 MINQUE(1)处理双列杂交数据 /73
 - 4.7.1 矩阵模型(全双列交配设计 Griffing III) /73
 - 4.7.2 平衡不平衡、规则不规则全双交配试验数据处理程序 /73
 - 4.7.3 不平衡的半双列交配设计单点试验(着重研究母本效应的大小) /75
 - 4.7.4 不平衡、不规则的半双列杂交试验数据的处理(不考虑父母效应) /76
- 4.8 两水平杂交试验不规则、不平衡试验资料的处理程序 /78
 - 4.8.1 试验简介 /78
 - 4.8.2 两水平杂交试验资料的处理技术 /78
 - 4.8.3 方差分析基础上的 MINQUE(1)法 /78
- 4.9 遗传相关和选择指数程序 /80
 - 4.9.1 同一试验设计的数据协方差分析程序 /80
 - 4.9.2 方差-协方差矩阵建立 /86
 - 4.9.3 相关系数的计算 /87
 - 4.9.4 计算选择指数 /88
 - 4.9.5 通径分析程序 /88
- 4.10 亲本育种值和子代个体育种值 /89
 - 4.10.1 单点单亲家系自由授粉子代试验资料的育种值分析程序 /89
 - 4.10.2 多点单亲家系自由授粉子代试验资料的亲本育种值分析程序 /94
 - 4.10.3 林木全同胞子代试验资料的个体育种值 /95
- 4.11 品种稳定性分析法(几何法) /107
 - 4.11.1 多点试验的数据采集 /107
 - 4.11.2 参试品种的均值、变异系数和成活率计算程序 /107
 - 4.11.3 参试品种材积的稳定性分析程序 /110
 - 4.11.4 参试品种木材比重的稳定性分析程序 /113
- 4.12 其他 /118
 - 4.12.1 林木田间试验中多地点区试不平衡资料的分析方法 /118
 - 4.12.2 关于多年度试验数据的统计处理方法 /118
 - 4.12.3 林木多父本杂交不平衡试验数据的分析技术 /119

第5章 试验模型选择、数据采集和不平衡数据分析技巧 /120

- 5.1 单因素随机试验设计不平衡数据的方差分析法数据输入格式 /120
 - 5.1.1 单因素类内观察值不等的方差分析线性模型 /120
 - 5.1.2 不平衡单因素完全随机试验的典型资料 /120
- 5.2 单因素随机区组试验设计的线性模型和不平衡数据的采集格式 /120

- 5.2.1 转化后的线性模型 /120
- 5.2.2 不平衡单因素随机区组试验的典型资料 /121
- 5.3 两因素有众多重复次数且不平衡模型分析时数据采集格式 /122
 - 5.3.1 不平衡数据的统计分析模型 /122
 - 5.3.2 不平衡数据的典型资料 /122
- 5.4 三阶平衡不平衡巢式设计的线性模型和数据采集格式 /123
 - 5.4.1 转化后的线性模型 /123
 - 5.4.2 转化后的数据采集格式 /123
- 5.5 两因素随机区组设计线性模型及不平衡数据的采集格式 /124
 - 5.5.1 转化后的线性模型 /124
 - 5.5.2 两因素随区组试验不平衡资料的采集格式 /124
- 5.6 四阶平衡不平衡、规则不规则巢式设计的线性模型及数据采集格式 /125
 - 5.6.1 转化后的线性模型 /125
 - 5.6.2 转化后不平衡、不规则试验数据的采集 /126
- 5.7 林木双列杂交试验中的不平衡、规则不规则的模型与试验数据格式 /128
 - 5.7.1 不平衡、规则不规则双列杂交的线性模型 /128
 - 5.7.2 线性模型的矩阵表达式 /128
 - 5.7.3 不平衡、规则不规则双列杂交数据格式的采集 /128
- 5.8 关于试验数据不平衡时性状间的协方差分析的数据采集 /129
 - 5.8.1 单因素完全随机区组试验设计不平衡资料的协方差分析的数据采集 /129
 - 5.8.2 转化后的协方差线性模型 /130
- 5.9 林木两水平杂交试验数据处理的模型选择和分析技术 /130
 - 5.9.1 林木两水平杂交试验的模型选择 /130
 - 5.9.2 林木两水平杂交试验数据处理的分析技术 /131
- 5.10 品种多地点区域化试验资料的分析技术 /132
- 5.11 林木遗传育种试验中非平衡试验资料的处理技术 /132
 - 5.11.1 多地点的林木田间子代试验资料处理理论与技术 /132
 - 5.11.2 综合数据处理的线性模型 /133
 - 5.11.3 关于多年度试验数据的统计处理技术 /133
- 5.12 试验数据不平衡时性状间的遗传相关分析技术 /134
 - 5.12.1 相关系数的计算 /134
 - 5.12.2 单因素完全随机区组试验设计不平衡时的指数选择理论 /134
 - 5.12.3 约束选择指数 /135
- 5.13 运用最佳线性预测理论评价育种值的分析技术 /136
 - 5.13.1 单点单亲家系自由授粉子代试验资料的育种值分析 /136
 - 5.13.2 亲本育种值的评价 /136
 - 5.13.3 个体育种值的预测 /136
 - 5.13.4 多点单亲家系自由授粉子代试验资料的育种值分析 /137

- 5.14 林木全同胞子代试验资料的个体育种值的估算技术 /137
 - 5.14.1 林木双列杂交设计的非平衡试验资料的遗传方差分量估计 /137
 - 5.14.2 林木杂交设计非平衡试验资料各因子效应的 BLUP 估计 /138
 - 5.14.3 林木双列杂交子代林中的个体育种值估计 /138
- 5.15 多点试验不平衡数据时林木品种高产性稳定性评价技术 /138
 - 5.15.1 林木品种稳定性分析的基本资料 /138
 - 5.15.2 林木品种的稳定性分析——多面体的体积计算公式 /139
 - 5.15.3 林木优良品种稳定性分析的步骤 /140
- 5.16 林木多父本杂交非平衡试验数据的统计分析技术 /140
 - 5.16.1 田间试验和非平衡试验数据的统计分析方法和技术 /140
 - 5.16.2 分组随机区组试验的非平衡资料的分析方法 /140

参考文献 /142

附 录 /144

- 附录 1 三阶不平衡不规则的巢式设计试验数据分析原理及 M 程序 /144
 - 1.1 各参试因子效应的平方和 /144
 - 1.2 该模型的期望方差结构 /144
 - 1.3 各因子的自由度 /145
 - 1.4 各因子的方差分量系数 /145
 - 1.5 参试因子的方差分量和遗传力 /146
 - 1.6 林分/家系/个体三水平不平衡、不规则试验数据处理 M 程序 /147
- 附录 2 析因设计不平衡、规则不规则数据的约简模型分析法 /152
 - 2.1 转化后的线性模型 /152
 - 2.2 参试因子效应离差平方和的分解 /152
 - 2.3 各参试因子效应的平方和 /153
 - 2.4 期望均方结构 /153
 - 2.5 自由度的分解 /155
 - 2.6 期望均方结构 /156
 - 2.7 各因子效应方差分量的求解遗传力的计算 /156
 - 2.8 各因子的 F 检验 /157
 - 2.9 主效因子的多重对比 /160
 - 2.10 析因设计的约简模型的 M 程序 /161

转化分析法的统计学基础

1.1 导言

林木田间试验是林木遗传育种的中心环节。由于外在环境的影响以及株间竞争等因素的作用,随着试验时间的推移,试验林的部分植株会死亡,原先造林时平衡的试验设计失去了均衡性,精心设计的试验获得非平衡的试验资料。此时采用传统的方差分析或现代线性模型理论来处理这些非平衡的试验数据,经常获得负的方差分量(线性模型中有的方法还常常出现迭代不收敛等问题)。负的方差分量,这既无生物学意义,又无从给出其数学解释,只能说明这些方法有问题。2008 年,齐明通过对试验设计及其原理,进行反复钻研,从试验设计的角度,提出了一个转化理论:其一,对于一些规则的正交试验设计类型,其非平衡的试验资料可以转化为随机区组设计试验来进行数据处理;其二,对于重复内环境一致的区组,多株小区可以转化为单株小区,这样原先 b 个区组, k 株小区试验,便变成了 bk 个重复,单株小区的试验,来进行统计分析;其三,对于更复杂的模型,经分析发现非正态分布因子获得负的方差分量时,可删除非正态参试验因子,重构线性模型进行统计分析。这就是转化分析法基本观点。这样对试验数据进行统计处理,可以解决参因子负的方差分量等问题。但是转化分析法是从试验设计的角度提出来的。Abderrahmaxe Achouch 和朱军等人(2002)采用蒙特卡罗模拟(Monte Carlo Simulation),进行比较同一试验资料,不同统计分析模式的效益。国内外不少学者采用蒙特卡罗模拟研究方差分量估计法的优劣。本研究借助于 Monte Carlo 模拟法,从统计学的角度来评价林木转化分析法的分析效果,为林木转化分析法的应用提供科学依据。

1.2 模型与方法

1.2.1 平衡试验设计及其统计分析模型

考虑计算工作量的大小和研究结果的普遍性,本研究选取单因素随机区组试验,RCB 设计,作为研究对象,由于在林木子代试验中,通常采用 4 株、5 株、6 株小区(偶尔也采用 8 株和 10 株小区)进行试验,故 Monte Carlo 模拟法共选用如下五个试验:

- 其一,40 个半同胞家系,10 个区组,4 株小区参试;
- 其二,40 个半同胞家系,8 个区组,5 株小区参试;
- 其三,40 个半同胞家系,8 个区组,6 株小区参试;
- 其四,40 个半同胞家系,6 个区组,8 株小区参试;

其五, 40 个半同胞家系, 4 个区组, 10 株小区参试;

选用 40 个半同胞家系, 是为了反映研究性状内在的遗传变异幅度。

以单株观察值作为统计分析单位, 有如下线性模型:

$$y_{ijk} = u + f_i + b_j + (fb)_{ij} + e_{ijk} \quad (1-1)$$

式中: $i = 1 \rightarrow 40$; $j = 1 \rightarrow 4 \sim 10$; $k = 4 \sim 10$; y_{ijk} 为第 i 个处理在第 j 个区组的第 k 个观察值; u 为群体平均值(为 60); f_i 为第 i 个半同胞家系的效应值, $f_i \sim N(0, \sigma_f^2)$; b_j 为第 j 个区组的效应值; $b_j \sim N(0, \sigma_b^2)$; e_{ijk} 为随机误差, $e_{ijk} \sim N(0, \sigma_e^2)$; $(fb)_{ij}$ 为家系与区组重复间的交互作用; 由于存在缺区或缺株, 绘散点图表明 $(fb)_{ij}$ 遵从二项分布。

1.2.2 平衡试验数据的产生

根据林木遗传育种的经验, 均假定这些因子服从正态分布, 指定机误的方差 V_e 为 280, 区组因子的方差 V_b 为 15, 半同胞家系的方差 V_f 为 20。在未对多株小区进行转化时, 非平衡资料的原试验设计中, 处理与区组重复间的交互作用因子不遵循正态分布, 绘散点图表明 $(fb)_{ij}$ 遵从二项分布, 故按二项分布产生其效应值。群体平均值为 60。采用 MATLAB7. X 语言编写的产生程序, 以获得试验的平衡资料。一个试验, 一次模拟 100 套数据, 并保存在 Excel 中。

1.2.3 非平衡试验数据的产生及统计分析

根据林木遗传育种中的经验, 造林存活率为 75%~95%, 但是杉木试验林通常的造林存活率为 83% 左右。假设为随机死亡, 存活率为 83%, 使用 MATLAB 语言中的删除语句, 获得非平衡试验数据。

式(1-1)不平衡状态下的方差分析原理见参考文献(齐明, 2009)。其方差分析方法亦采用 MATLAB7. X 语言, 编写试验数据分析程序, 以便获得未转化的每套试验资料包括方差分量在内的诸多参数。参见本书的第 4 章有关章节。

1.2.4 试验数据的转化分析法

每套资料在完成以上分析后, 将多株小区转化为单株小区后, 转化后的线性模型:

$$y_{ijk} = u + f_i + b_j + e_{ijk} \quad (1-2)$$

式中: $i = 1 \rightarrow 40$; $j = 1 \rightarrow 40 \sim 48$; $k = 1$ 或 0; u 为群体平均效应; 其他各因子的意义与式(1-1)中相同。

在随机模型条件下, 各参试因子都是随机因子。于是有:

$$f_i \sim N(0, \sigma_f^2); b_j \sim N(0, \sigma_b^2); e_{ijk} \sim N(0, \sigma_e^2)$$

方差分析原理参见有关文献(齐明, 2009)。分析方法亦是采用 MATLAB7. X 语言编写的程序进行, 分析目的同样是为了获得机误的方差 V_e , 区组因子的方差 V_b , 半同胞家系的方差 V_f , 半同胞家系遗传力 h_f^2 , 家系内单株遗传力 h_i^2 。

1.2.5 Monte Carlo 模拟数据分析结果的评价

根据 100 次 Monte Carlo 模拟结果, 计算出若干参数的估计值, 参数的偏差(偏差 = 估计值 - 参数真值)、偏差的显著性检验: 如果偏差的绝对值被其估计值除得的商, 大于