林业科学研究 2004,17(3):279~283
Forest Research

文章编号:1001-1498(2004)03-0279-05

用 Mixed 和 Nlmixed 过程建立混合生长模型

李永慈[1], 唐守正[2]
(1. 北京林业大学资源与环境学院,北京 100083; 2. 中国林业科学研究院资源信息研究所,北京 100091)

摘要:本文用 5 块不同密度样地的树高生长资料,根据线性和非线性混合模型理论,利用 SAS 的 Mixed 过程和 Nlmixed 过程,分别拟合线性混合模型和非线性混合树高生长模型。根据预测值和固定效应同时绘制出不同密度下的高生长曲线和平均高生长曲线,充分显示了混合模型的优势,即它可以同时反映总体的平均变化趋势和个体之间的差异。

关键词:混合模型;随机效应;固定效应;生长模型
中图分类号:S758.5 文献标识码:A

混合模型是近代发展起来的新的统计方法,很多通常的统计模型可以看作混合模型的特例。在混合模型中,固定效应是描述总体平均变化趋势,随机效应是描述从总体中抽取的个体。混合模型主要用来研究分组数据中因变量和自变量的关系,它的优势在于它的灵活性,它既可以反映总体的平均变化趋势,又可以反映个体之间的差异。目前国外统计分析软件 SAS 系统和 S-PLUS 的最新版已经先后推出线性和非线性混合模型模块,SPSS 系统的最新版也提供了线性混合模型的功能,但是还没有提供非线性混合模型的功能。混合模型在农业、生态、经济、制造等行业有着广泛的应用,早在 20 世纪 70 年代国外就有人开始用混合模型研究生长模型,然而国内介绍混合模型理论和软件应用的文献较少,用混合模型研究生长模型在国内还是一个空白。本文的目的在于介绍如何应用 SAS 的 Mixed 和 Nlmixed 过程建立混合生长模型。

1 混合模型

混合模型分为线性混合模型和非线性混合模型。线性混合模型比广义一元线性模型多了随机参数部分,当随机参数的设计矩阵恒等于零时,它就退化为广义一元线性模型。非线性混合模型中,部分或全部的固定效应和随机效应在模型中与因变量呈非线性关系,因此它可以看成线性混合模型的推广;同时它又可以看成非线性回归模型的推广,因为它可以看成是将随机效应引入非线性回归模型而得到的。

1.1 线性混合模型

给出两水平的线性混合模型,这个模型很容易推广到多水平的情形。

因变量 y_{ij} 是一个 n_{ij} 维的向量,可以表示为 $y_{ij} = X_{ij}\beta + Z_{i,j}b_i + Z_{ij}b_{ij} + \varepsilon_{ij}$, $i = 1, \cdots, M$, $j = 1$,

收稿日期:2003-06-18
基金项目:948 项目"数字林业"关键技术引进,项目编号:2001-13
作者简介:李永慈(1965—),女,北京林业大学资源与环境学院在读博士,副教授.

$\dots, M_iM$ 是第一水平的组数, M_i 是第一水平的第 i 组中第二水平的组数。 $b_i \sim N(0, G_1(\theta_{G_1}))$, $b_{ij} \sim N(0, G_2(\theta_{G_2}))$, $\varepsilon_{ij} \sim N(0, \sigma^2 I)$。$\beta$ 是 p 维的固定效应向量, b_i 是 q_1 维的随机效应向量, b_{ij} 是 q_2 维的随机效应向量, $X_{ij}(n_{ij} \times p)$ 和 $Z_{i,j}(n_{ij} \times q_1)$、$Z_{i,j}(n_{ij} \times q_2)$ 是固定效应和随机效应的设计矩阵, ε_{ij} 是 n_{ij} 维的组内误差向量。

对于不同的 i, b_i 是相互独立的; 对于不同的 i 或 j, b_{ij} 是相互独立的并且独立于 b_i, ε_{ij} 是相互独立的并且独立于 b_i 和 b_{ij}。因此 $y_{ij} \sim N(X_{ij}\beta, V_{ij}(\theta))$, $V_{ij}(\theta) = Z_{i,j}G_1(\theta_{G_1})Z_{i,j}^T + Z_{ij}G_2(\theta_{G_2})Z_{ij}^T + \sigma^2 I$。两水平线性混合模型的扩展将对 $\varepsilon_{ij} \sim N(0, \sigma^2 I)$ 的限制放松为 $\varepsilon_{ij} \sim N(0, \sigma^2 \Lambda_{ij})$。

1.2 非线性混合模型

给出两水平的非线性混合模型, 这个模型很容易推广到多水平的情形。

$y_{ijk} = f(\phi_{ijk}, v_{ijk}) + \varepsilon_{ijk}$, $i = 1, \dots, M$, $j = 1, \dots, M_i$, $k = 1, \dots, n_{ij}$, M 是第一水平下的组数, M_i 是第一水平下的第 i 组中第二水平的组数, n_{ij} 是第一水平下的第 i 组中第二水平下第 j 组的观测数。f 是分组参数 ϕ_{ijk} 和 v_{ijk} 的实值可微函数, f 至少是 ϕ_{ijk} 中一个参数的非线性函数。其中, $\phi_{ijk} = A_{ijk}\beta + B_{i,jk}b_i + B_{ijk}b_{ij}$, $b_i \sim N(0, G_1(\theta_G))$, $b_{ij} \sim N(0, G_2(\theta_G))$, $\varepsilon_{ijk} \sim N(0, \sigma^2)$。用矩阵形式可以表示为: $y_{ij} = f_{ij}(\phi_{ij}, v_{ij}) + \varepsilon_{ij}$, $\phi_{ij} = A_{ij}\beta + B_{i,j}b_i + B_{ij}b_{ij}$, $i = 1, \dots, M$, $j = 1, \dots, M_i$, 这里 $y_{ij} = \begin{bmatrix} y_{ij1} \\ \vdots \\ y_{ijn_{ij}} \end{bmatrix}$, $\phi_{ij} = \begin{bmatrix} \phi_{ij1} \\ \vdots \\ \phi_{ijn_{ij}} \end{bmatrix}$, $f_{ij}(\phi_{ij}, v_{ij}) = \begin{bmatrix} f(\phi_{ij}, v_{ij1}) \\ \vdots \\ f(\phi_{ijn_{ij}}, v_{ijn_{ij}}) \end{bmatrix}$, $v_{ij} = \begin{bmatrix} v_{ij1} \\ \vdots \\ v_{ijn_{ij}} \end{bmatrix}$, $A_{ij} = \begin{bmatrix} A_{ij1} \\ \vdots \\ A_{ijn_{ij}} \end{bmatrix}$, $B_{ij} = \begin{bmatrix} B_{ij1} \\ \vdots \\ B_{ijn_{ij}} \end{bmatrix}$。两水平非线性混合模型的扩展将对 $\varepsilon_{ij} \sim N(0, \sigma^2)$ 的限制放松为 $\varepsilon_{ij} \sim N(0, \sigma^2 \Lambda_{ij})$。

2 实例

2.1 数据来源

江西省大岗山实验局 5 块初植密度不同的杉木(*Cunninghamia lanceolata* (Lamb.) Hook.), 每个密度下选择 3 块样地, 分别在 5、6、7、8、9、10、11、12、14、16 年生测量优势树高。

表 **1** 同密度下杉木优势高

年龄/a	样地														
	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3	4-1	4-2	4-3	5-1	5-2	5-3
5	5.8	5.4	5.2	5.6	5.5	5.5	5.5	5.1	5.5	5.0	5.0	5.2	5.3	4.8	5.5
6	7.2	6.7	6.9	7.0	6.6	6.6	6.8	6.1	6.4	5.8	5.9	6.0	6.0	6.0	6.3
7	8.1	7.6	8.4	8.1	7.3	7.6	8.3	7.1	7.2	6.7	6.5	6.9	7.1	6.9	7.8
8	9.4	8.6	9.7	9.0	8.0	9.0	8.9	8.2	8.8	7.5	7.5	7.7	8.3	8.0	8.4
9	10.6	9.4	10.7	10.0	9.0	10.2	10.1	8.7	9.8	8.2	8.3	8.5	9.0	9.1	9.3
10	11.8	9.9	11.9	11.2	9.8	11.4	11.0	9.5	10.3	8.8	8.8	9.4	9.9	9.8	9.8
11	12.8	10.8	13.1	11.6	10.4	12.0	11.6	10.2	11.2	9.4	9.5	10.0	10.4	10.4	10.7
12	12.7	11.0	13.3	12.2	10.8	13.0	12.0	10.7	11.3	9.6	9.8	10.2	10.7	10.8	10.9
14	14.0	11.7	14.4	13.4	11.6	14.0	12.6	11.6	12.2	10.4	10.7	11.2	11.4	12.2	12.2
16	15.0	12.3	15.7	14.3	12.5	15.2	13.6	13.0	12.8	12.0	11.8	11.8	11.8	12.8	12.8

2.2　线性混合树高生长模型

林业上常用的苏玛克模型 $y=\beta_1\exp(\beta_2(1/20-1/x))$ 的对数形式是线性的, $\ln(y)=\ln(\beta_1)+\beta_2(1/20-1/x)$, 用 y 来记 $\ln(y)$, 用 β_1 来记 $\ln(\beta_1)$, 用 x 来记 $1/20-1/x$, 得 $y=\beta_1+\beta_2 x$。初植密度是要考虑的一个水平, 因此这是一个单水平的线性混合模型。$y_i=(\beta_1+b_{1i})+(\beta_2+b_{2i})x_i+\varepsilon_i$, $i=1,\cdots,5$。其中 b_{1i}、b_{2i} 是随机效应, 记 $b_i=\begin{bmatrix}b_{1i}\\ b_{2i}\end{bmatrix}$, 假定 $b_i\sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}sb_1 & s_{12}\\ s_{12} & sb_2\end{bmatrix}\right)$, $\varepsilon\sim N(0,\sigma^2 I)$ 对于每个组的 $X_i(30\times2)$ 与 $Z_i(30\times2)$ 相同, 都等于 $\begin{bmatrix}1 & -0.15\\ 1 & -0.15\\ 1 & -0.15\\ 1 & -0.1167\\ \vdots & \\ 1 & -0.0125\end{bmatrix}$。用 SAS 的 Mixed 过程拟合这个线性混合模型。

```
proc mixed data= sumake1;
class plot;
model y= x/s outp= a;
random intercept x/type= un subject= plot s;
run;
```

估计结果为: $\beta_1=2.6452$, $\beta_2=6.6558$, $b_{11}=0.07921$, $b_{21}=-0.08503$, $b_{12}-0.07261$, $b_{22}=0.4368$, $b_{13}=0.07751$, $b_{23}=1.1968$, $b_{14}=-0.1199$, $b_{24}=-0.2215$, $b_{15}=-0.1094$, $b_{25}=-1.3271$, $sb_1=0.008364$, $sb_2=0.04032$, $s_{12}=0.02521$, $\sigma^2=0.003053$。由预测结果做出的 5 种不同密度下的高生长曲线和由固定效应所得的平均生长曲线如图 1。

2.3　非线性混合树高生长模型

非线性树高生长模型选择林业上常用的 Logistic 模型 $y=\beta_1/(1+\beta_2\exp(-\beta_3 x))$。初植密度是需要考虑的一个水平, 因此这是一个单水平的非线性混合模型。$y_{ij}=(\beta_1+b_{1i})/(1+(\beta_2+b_{2i})\exp(-(\beta_3+b_{3i})x_{ij}))+\varepsilon_{ij}$, $i=1,\cdots,5$, $j=1,\cdots,10$, 其中 b_{1i}、b_{2i}、b_{3i} 是随机效应, 记 $b_i=\begin{bmatrix}b_{1i}\\ b_{2i}\\ b_{3i}\end{bmatrix}$, 假定 $b_i\sim N\left(\begin{bmatrix}0\\0\\0\end{bmatrix},\begin{bmatrix}sb_1 & s_{12} & s_{13}\\ s_{12} & sb_2 & s_{23}\\ s_{13} & s_{23} & sb_3\end{bmatrix}\right)$, $\varepsilon_{ij}\sim N(0,s)$。用 SAS 的 Nlmixed 过程拟合这个非线性混合模型。

```
proc nlmixed data= modellyc;
parms beta1= 13.8  beta2= 6.7  beta3= 0.3  sb1= 0.9264  sb2= 1.2684
sb3= 0.000966  s12= 0  s13= 0  s23= 0  s= 0.82;
fm1= exp(-(beta3+ b3)×x);
fm2= 1+ (beta2+ b2)×fm1;
mea= (beta1+ b1)/fm2;
```

```
model Ho2~ normal(mea, s);
random b1 b2 b3~ normal([0, 0, 0], [sb1, s12, sb2, s13, s23, sb3]) subject= plot;
predict mea out= p;
```

估计结果为: $\beta_1 = 13.794\ 2$, $\beta_2 = 6.715\ 5$, $\beta_3 = 0.297\ 8$, $sb_1 = 0.914\ 6$, $sb_2 = 1.178\ 9$, $sb_3 = 0.000\ 273$, $sb_{12} = 0.071\ 12$, $s_{13} = 0.012\ 70$, $s_{23} = 0.011\ 66$, $s = 0.339\ 5$。由预测结果做出的5种不同密度下的高生长曲线和由固定效应所得的平均生长曲线如图2。

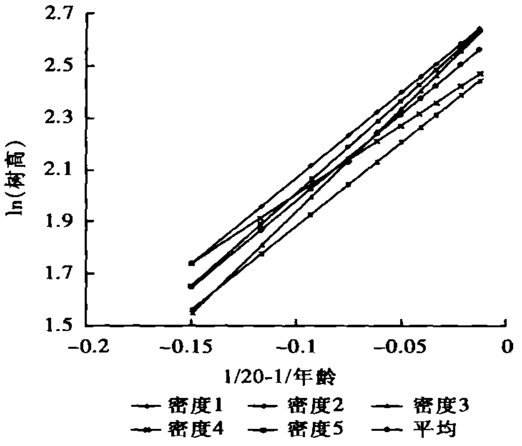

图1 变换后苏马克混合模型的拟合图

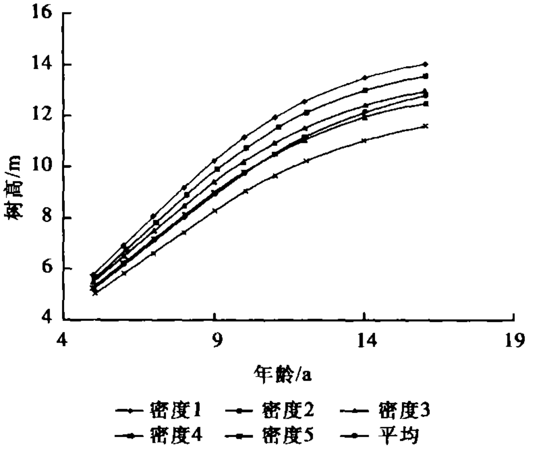

图2 非线性Logistic混合模型的拟合图

3 结论

非线性Logistic混合模型很好地拟合了原始树高生长过程,变换后的苏马克混合模型很好地拟合了变换后的树高生长过程。线性混合模型允许数据之间是相关的和不等方差的,非线性混合模型允许用户为数据指定多种分布形式,因此混合模型有着以往其它模型不可比拟的优势,就是它既可以反映总体的平均变化趋势,又可以提供数据方差、协方差等多种信息来反映个体之间的差异。

参考文献:

[1] Tang S Z, Meng F R. Analyzing parameters of grouth and yield models for Chinese fir provenances with a linear mixed approach[J]. Silvae Genetica, 2001, 50: 140~145

[2] José Pinheiro, Douglas M Bates. Mixed-Effects Models in S and S-PLUS[M]. Springer Verlag New York, Inc, 2000

[3] SAS Institute Inc. SAS/STAT User's Guide Version 8[M]. SAS Institute Inc, Cary, NC, USA, 1999

[4] Daniel B H, Robert L B. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models[J]. Forest Science, 2001, 47(3): 311~321

[5] 唐守正, 李勇. 生物数学模型的统计学基础[M]. 北京: 科学出版社, 2002

Establishment of Tree Height Growth Model Based on Mixed and Nlmixed of SAS

LI Yong-ci[1], *TANG Shou-zheng*[2]

(1. College of Forest Resources and Environment, Beijing Forestry University, Beijing 100083, China;

2. Research Institute of Forest Resource Information Techniques, CAF, Beijing 100091, China)

Abstract: Based on the data of height growth in 5 plots with different densities, mixed and nlmixed of SAS were used to fit the linear and non-linear mixed models of tree height growth respectively. The height growth curve and mean height growth curve under different densities were drawn according to calculated value and fixed effect value, which showed the advantages of the mixed model, i. e. it could reflect the mean variation trend on the 5 plots and the differences among individuals at the same time.

Key words: mixed model; random effect; fixed effect; growth model