

Tab 5 Some additional files

name	type	description
RELWDAT PRN	TXT	— Renumbered data file
RELWPED PRN	TXT	— Renumbered Pedigree file
RENUM PAR	TXT	— Parameter file for program RENUMAT
RENUM MSG	TXT	— Log file from renumbering Describe the levels of fixed and random effects after renum
RENUM PRN	TXT	— The details from renumbering Describe how effects are combined and replications of each effects
BLUP PAR	TXT	— Parameter files for REML and BLUP analysis
SOLUTIONS	TXT	— BLUP solutions file
SOLUTIONS_VCE	TXT	— This file keeps variance estimation if performed

questions, constructive criticism and debug reports are invited. Please email to monchai@kku.ac.th or gnacy@uga.edu

BLUPF90-PSPAK has been made available on CD. However, the updated version is available at BLUPF90 homepage at KKU <http://agserver.kku.ac.th/monchai>. The complete package provides program files, manuals and examples. Online registration is requested for further breeding and genetic group connection and to update version information. The registration page for BLUPF90-PSPAK is also available at BLUPF90 homepage.

References

[1] CHEN P, BASST J, Mahdy JW, et al. Genetic

correlations between lean growth and litter traits in US Yorkshire, Duroc, Hampshire, and Landrace pigs J. J Anim Sci 2003 84: 1700—1705.

[2] NSIF. Guidelines for Uniform Swine Improvement Programs J. Washington, DC: USDA, 1997.

[3] NSR. STAGES National Genetic Evaluation Trait Leader List J. N. National Swine Registry, 2000.

[4] LYNCH M, WALSH B. Genetics and Analysis of Quantitative Traits J. CA: Sinauer Association, 1998.

(Edited by ZHOU Zhi-hong)

译文

动物育种中解决混合模型复杂计算问题的软件应用

Monchai DAUNGJINDA

(Department of Animal Science Faculty of Agriculture, Khon Kaen University, Thailand 40002)

摘要: 在动物育种和遗传选择的改良计划中应用的软件有很多个。本文描述了其中最常用的软件, 并提出了有关软件的改进建议。本文作者供职于泰国的 Khon Kaen 大学, 可以免费提供其新开发的软件。需要该软件的请与作者联系: monchai@kku.ac.th

关键词: 混合模型; 计算软件; 动物育种

最近, 有数个软件包可以用于动物育种的混合模型分析。对于遗传评估和选择方案来说, 进行方差组分估计和遗传性能预测需要特殊的软件。虽然有些可以通过国际互联网下载, 可是, 多数这样的软件

包为商业软件或者用起来不方便. 表 1列出了最近可以在国际互联网上获得的软件包.

用于育种工作的多数软件是免费的(多数情况下为研究用),有的还提供源代码,不过,一般操作平台为 UNIX系统. 把这些程序用在 Windows平台则需要特定的编译器. 一般地,需要特定的准备数据文件和写出操作文件.

1 BLUPF90-P iSPAK

编写 BLUPF90-P iSPAK的主要目的是为猪遗传评估模型提供分析方法,而且可用于 PC机和 Windows操作系统. 有用户友好的图形接口. 利用 BLUPF90功能强大的特点, P iSPAK可以进行许多遗传评估方面的分析. P iSPAK可利用 REML原理进行

方差组分估计,利用线性混合模型进行 BLUP育种值分析,其中随机效应包含动物个体加性遗传和永久环境效应.

“BLUPF90-P iSPAK”可在 Microsoft EXCEL下运行. 进行遗传评估时,它有极好的接口,通过 5个步骤可以产生 VCE和 BLUP报告. 分析只需要提供数据文件和系谱文件. 它可以自动地对动物个体号进行重新编码,对模型中各效应的水平数进行计数. 分析繁殖性能时可以有配偶效应(或与配公猪效应),分析生长性状时可以包括窝效应,这些效应在数据文件中可以字母格式编写. 方差组分分析或 BLUP分析结束后, P iSPAK能够以个体原始 ID号输出育种值(BV)报告. 使用过程中用户可以分别准备数据文件和系谱文件,不过,存放这些文件的路径要在分析

表 1 最近可用于动物育种中混合模型分析的软件包

名称	详细说明
ASREML	专门用于大型数据集,拟合混合模型的统计软件包. 包括单性状、多性状和重复度量性状和空间分析等的注释代码. http://uncronopio.org/asrem/homepage
DFREML	可用于动物育种的方差组分估计软件. 由 Karin Meyer开发的、利用非求导 REML进行统计分析. http://agbu.unc.edu.au/~kmeyer/dfrem1.html
BLUPF90 / REMLF90 / GIBBSF90	可用于动物遗传评估的软件包. 可以通过 REML和贝叶斯分析(Gibbs抽样)进行 BLUP育种值和方差组分估计. http://nce.ads.uga.edu/~ignacy/Programs.html
PIGBLUP	用于猪育种的 BLUP软件. http://abrj.unc.edu.au
VCE/PEST	VCE是由 Groeneveld开发的进行方差组分估计的软件. 还包括用 PEST估计育种值的文件. http://www.tzy.fal.de/~eg/
MTDFREML	Keith Boldman和 Dale Van Vleck开发的程序. Lisa Kriese和 Curt Van Tassel还提供了调试说明. http://www.arsusda.gov/curvym/dfrem1.html

前说明. 带有估计准确性的 BV报告最多只能有 4个性状.

“BLUPF90”及相关的程序是由 Ignacy Misztal博士实验室开发的,其目的是提供强大的、用于动物育种中混合模型分析的软件. 详细文件可以在 <http://nce.ads.uga.edu/~ignacy>中找到. 多数程序是用 Fortran 90编写的,并且有一个线性接口. “P iSPAK”是一组从“PCPAK”软件中发展出来的程序,在 Windows下有图形使用界面,使用简单. P iSPAK是专门用于猪遗传评估的程序, BLUPF90软件中的部分程序也可以使用,它们是:

- BLUPF90 用 PCG方法进行 BLUP分析.
- REMLF90 利用 EM-REML进行方差组分估计.

AIREMLF90 利用 AIREML算法进行方差组分估计.

RENUMMAT 重新编号程序,可以产生数据文件和加性系谱文件,产生的新文件中动物个体 ID已经排序.

ACCF90 可以计算直接效应、母体效应和多性状模型 BLUP解的近似精确度.

所有的程序都是在 Microsoft Windows操作系统下用 Visual Fortran Version 5.1版本分别编译的. 用户可以在 MS-DOS提示符或者在 P iSPAK下分别应用. 这样,用户可以在 MS-EXCEL中用 Visual BASIC编写的界面上,通过鼠标点击的方式获得所有的参数文件和 BLUP报告.

2 规格

PigPAK需要在 Windows95/98/ME/XP/2000环境下安装,运行环境为 Excel98/2000/XP 它需要至少 32 MB的内存和 5 MB的硬盘空间以储存程序。

利用 BLUP90和 REML90可以进行单性状和多性状的 BLUP分析和方差组分分析,如动物模型、重复力模型、母体效应模型、显性效应模型和随机回归模型等,可以考虑缺失数据,不同的性状可以是不同的模型。AREML90目前可能还不支持某些模型和特殊的数据结构。ACCF90是重复力模型和母体效应模型的近似分析程序。

利用 PigPAK接口时,如果选择单性状分析,则只能用于简单动物模型和 PE动物模型。多性状分析最多可以有 4个性状。只在单性状和多性状模型中提供了 BLUP估计准确性的结果。在这一界面下,用 REML或 AREML估计方差组分仍然是可行的。

3 模型描述

下面是用于分析的模型的详细说明。

3.1 基本动物模型

PigPAK包括一种最基本的动物模型,利用它可以对数据集中的个体和系谱中的个体进行分析,这时考虑了所有的亲缘关系。其他的效应,无论是固定效应还是随机效应,都可以将混合模型技术包括在模型中。模型中固定效应可以为交叉分组变量和协变量。在分析时可以有组合效应,如场一年一季效应,因此,这时不需要有进一步的数据准备。正常情况下,正确利用 PigPack的线性混合模型进行分析时,所有性状都应该是连续的,而不应该是分类变量。对于每个动物个体只有一条记录的性状模型,如生长性状,可以用基本模型分析如下:

y= Xβ + Za+ε V [a] = [Aσ_a^2 0 ; 0 b_c^2]

上式中, y为反应变量向量,β 为固定效应向量, a为随机加性效应向量, ε为随机残差向量, X和 Z分别为与固定效应和随机效应有关的结构矩阵, A为分子亲缘相关矩阵, σ_a^2为加性遗传方差, σ_c^2为残差效应方差。

进行 BLUP分析时, Hendersor的混合模型方程 (MME)可以重新写成:

[XX' X'Z ; Z'X Z'Z+α A] [β] = [X'y ; Z'y]

这里: α = σ_c^2 / σ_a^2

3.2 重复力模型

如果有多胎数据,则需要考虑个体永久环境效应。如果公猪或母猪窝效应可知,也应作为随机效应考虑,见 Cher等^[1]。一般地,在遗传评估中,通常把永久环境效应作为独立的随机效应。分析用的模型为:

y= Xβ + Za+ Wc+ε V [a ; c ; ε] = [Aσ_a^2 0 0 ; 0 b_c^2 0 ; 0 0 b_e^2]

上式中, y为反应变量向量,β 为固定效应向量, a为随机加性效应向量, c为随机永久环境效应向量, ε 为随机残差向量, X W和 Z为与随机效应和固定效应相关的结构矩阵, A为分子亲缘相关矩阵, σ_a^2为加性遗传方差, σ_c^2为永久环境效应方差, σ_e^2为残差效应方差。

进行 BLUP分析时, MME可以写成:

[X'X X'Z X'W ; Z'X Z'Z+α A^{-1} Z'W ; W'X W'Z W'W+ν I] [β ; a ; c] = [X'y ; Z'y ; W'y] , 这里: α = σ_c^2 / σ_a^2 , ν = σ_e^2 / σ_c^2

3.3 多性状模型

利用 PigPAK也可以进行多性状分析。性状间遗传相关的估计和多性状 BLUP分析可以利用国家猪改良联合会 (NSF)^[2] 和国家猪注册协会 (NSR)^[3] 的模型进行。不过,如果想利用图形界面,性状不能超过 4个。性状超过 4个时,需要利用菜单 “BLUP> Use old parameters”进行参数编辑。只有一个记录或重复记录的性状都可以进行多性状分析,这时的模型不同。每个性状只有一条记录的两性状模型如下:

[y_1 ; y_2] = [X_1 0 ; 0 X_2] [β_1 ; β_2] + [Z_1 0 ; 0 Z_2] [a_1 ; a_2] + [ε_1 ; ε_2] , V [a_1 ; a_2] = [G ⊗ A 0 ; 0 R ⊗ I]

式中, y_1, y_2 分别为性状 1和 2的反应变量向量; β_1, β_2 分别为固定效应向量; α_1, α_2 分别为随机加性遗传效应; ε_1, ε_2 分别为随机残差效应向量; X_1 W和 Z_1 为与与固定效应和随机效应相关的结构矩阵; A为分子亲缘相关矩阵; G为性状 1和 2的直接遗传方差—协方差矩阵; R为性状 1和 2的残差方差—协

方差矩阵.

3.4 抽样方差

如果在应用 PiSPAK时选择了 AREML,则可以计算每个方差组分的抽样方差. 因此, 根据 Lynch和 Walsh^[4]的方法, 遗传力的标准误可以如下计算:

$$\text{Var}\left(\frac{\sigma_a^2}{\sigma_t^2}\right) = \left(\frac{\sigma_a^2}{\sigma_t^2}\right)^2 \left[\frac{\text{Var}(\sigma_a^2)}{(\sigma_a^2)^2} + \frac{\text{Var}(\sigma_t^2)}{(\sigma_t^2)^2} + \frac{2 \text{cov}(\sigma_a^2, \sigma_t^2)}{(\sigma_a^2)(\sigma_t^2)} \right].$$

则: $SE(\hat{h}) = \sqrt{\text{Var}\left(\frac{\sigma_a^2}{\sigma_t^2}\right)}$

另外, 对于多性状分析来说, 遗传相关的标准误可以用下式近似计算:

$$\text{Var } r = (\hat{r}) = (r)^2 \left[\frac{\text{Var}(\sigma_i^2)}{4(\sigma_i^2)^2} + \frac{\text{Var}(\sigma_j^2)}{4(\sigma_j^2)^2} + \frac{\text{Var}(\sigma_{ij})}{4(\sigma_{ij})^2} + \frac{2 \text{cov}(\sigma_i^2, \sigma_j^2)}{4(\sigma_i^2)(\sigma_j^2)} + \frac{2 \text{cov}(\sigma_i^2, \sigma_{ij})}{2(\sigma_i^2)(\sigma_{ij})} + \frac{2 \text{cov}(\sigma_j^2, \sigma_{ij})}{2(\sigma_j^2)(\sigma_{ij})} \right]$$

于是:

$$SE(r) = \sqrt{\text{Var } r}.$$

3.5 程序的使用

程序安装后, 程序存储在主目录或者用户定义的目录中, 如 C:\BIUPF90-PP 还有两个子目录存放例子和帮助文件. 在各目录中可以发现表 2至表 4中列出的程序和文件.

表 2 主目录

名称	类型	描 述
PiSPAK .xls	XLS	— 用户图形主界面 — 产生 BLUP和 REML参数文件 — 产生 BLUP和 VCE报告, EXCEL格式 — 产生遗传趋势 — 计算测定日模型的 BV和 R.
BIUPF90 .EXE	PROG	— 计算 BLUP解
REMLF90 .EXE	PROG	— 利用 REML的 EM算法估计方差组分
AIREFMLF90 .EXE	PROG	— 利用 REML的 A算法估计方差组分
ACCF90 .EXE	PROG	— 计算 BLUP解的近似准确性
RENUMMAT .EXE	PROG	— 把数据文件和系谱文件中的数据重新排序, 结果为连续的数字.

表 3 例子目录

名称	类型	描 述
LW DAT .PRN	TXT	— 有重复记录的单性状分析和多性状分析的数据文件
LW PED .HRN	TXT	— 利用 LW DAT分析的系谱文件
LW DAT .HMT	TXT	— 用于 LW DAT .PRN程序的描述列数字格式

表 4 帮助目录

名称	类型	描 述
WHOSWHQ .TXT	TXT	— BIUPF90系列主要开发人员的授权声明
Manual-PP .PDF	PDF	— BIUPF90 PiSPAK手册

当利用 PiSPAK界面进行分析时, 所有的参数都要写入特定的 Excel工作表中保存. 它们将被 Visual Basic程序写成正确的参数文件格式, 用于 RENUM和 BIUPF90 调用这些 VB程序需要用批命令, 可以利用 Excel中特定的 VB函数来运行该批命令中的 EXE文件, 而不用关闭 Excel程序. 在程序运行过程中, 系谱和计算结果都读入 Excel工作表, 与原始的个体 ID对照. 同时, 如果需要, 利用绘图函数可以产

生遗传趋势图. PiSPAK中应用的数据和系谱文件必须是 ASCII或文本文件. 除了个体 ID公猪号、母猪号和配偶或窝效应可以用字母格式外, 所有数据必须是数字. 如果用 Excel产生, 则存储为 PRN文件 (用空格隔开的文本文件)比用 Tab分隔或逗号分隔更方便. 如果数据存储在不同的文件夹中, 所有的可执行程序会拷贝到数据文件夹, 以简化操作. 数据分析

完成后, 只会产生少数几个文件, 如果需要, 可以改变它们的名字.

PRN分析完成后有表 5 列出的额外文件, 它们在以后的分析中可能用得上.

假定最初的文件为 LW DAT PRN和 LW PE D

表 5 一些额外文件		
名称	类型	描 述
RELW DAT PRN	TXT	— 重新编号的数据文件
RELW PE D PRN	TXT	— 重新编号的系谱文件
RENUM PAR	TXT	— RENUM MAT程序使用的参数文件
RENUM MSG	TXT	— 重新排序后的 Log文件. 描述重新编号后固定效应和随机效应的水平数.
RENUM PRN	TXT	— 重新编号的详细信息. 描述效应是如何合并的, 每个效应的重复数.
BLUP PAR	TXT	— REML和 BLUP分析的参数文件.
SOLUTIONS	TXT	— BLUP解文件.
SOLUTIONS VCE	TXT	— 方差组分估计时, 该文件保存方差组分的估计结果.

缺省时的收敛标准为 1 d-08 不过, 使用人员可以通过菜单选择自己的收敛标准.

4 应用条件

BLUP90-P iPAK可以免费用于学术和研究目的, 条件是版权仍然属于作者. 应用 P iPAK或由该软件编程均必须在发表时予以说明. 对于商业或者有经费支持的应用, 可以和任何一位作者联系, 以取得作者的同意. 我们不能保证用户的正确使用和提供任何服务. 但是, 在特定的情况下, 有建设性的意见、调试报告等, 请利用下述的 email地址联系: mon-

chai@ kku. ac. th或 ignacy@ uga. edu

BLUP90-P iPAK目前已经做成 CD盘, 最新版本可以在 KKU的网页中下载, <http://agserver.kku.ac.th/monchai> 完整的软件包提供了程序文件、使用手册和例子. 如果要与该育种与遗传研究小组联系和获得最新版本的软件, 则需要网上注册. 在 BLUP90主页也可以完成 BLUP90-P iPAK的注册.

参考文献 (见原文)

【译者 张 豪; 责任编辑 周志红】