

混合线性模型下猪群间遗传联系的度量

王爱国¹, D Laloë², LR Schaeffer³

(1. 中国农业大学动物科学技术学院, 北京 100094; 2. Station de génétique quantitative et appliquée, INRA, 78352 Jouy-en-Josas cedex, France; 3. CGIL, University of Guelph, Guelph, Ontario, Canada N1G 2W1)

摘要: 采用关联指数(IC)和决定系数(CD)两种方法, 度量混合线性模型遗传评估下猪群体间的遗传关联性。结果表明, 加拿大安大略省的大约克夏猪、长白猪、杜洛克猪和汉普夏猪 4 个主要品种群体间具有良好的遗传联系。CD 法既组合了数据结构和信息量, 又考虑了预测误差方差和遗传变异性, 是一个选择判断遗传评估精度的好方法。

关键词: 混合线性模型, 遗传关联, 猪群体间

中图分类号: Q348, S828.2

文献标识码: A

文章编号: 0253-9772(2000)05-0295-03

Measures of Genetic Connectedness Between Herds in Swine Under Mixed Linear Models

WANG Ai-guo¹, D Laloë², LR Schaeffer³

(1. College of Animal Science and Technology, China Agricultural University, Beijing 100094, China; 2. Station de génétique quantitative et appliquée, INRA, 78352 Jouy-en-Josas cedex, France; 3. CGIL, University of Guelph, Guelph, Ontario, Canada N1G 2W1)

Abstract: Two criteria for the measures of genetic connectedness in mixed linear model of genetic evaluation are used: the degree of connectedness (IC) and the generalized coefficient of determination (CD). The results indicated that the data of four dominant swine breeds: Yorkshire, Landrace, Duroc and Hampshire in Ontario, Canada are well connected. The CD, which combines data structure and amount of information and also accounts for both prediction error variance and genetic variability, is a good method to select for judging the precision of a genetic evaluation.

Key words: mixed linear model, genetic connectedness, between herds in swine

在最佳线性无偏预测 (best Linear unbiased prediction, BLUP) 遗传评估中, 常常遇到的问题是精度 (precision) 和关联性 (connectedness), 它们对动物育种具有重要的意义。然而, 不关联 (disconnectedness) 是难以明确定义的, 有时定义为固定效应水平间缺乏遗传联系, 有时则定义为遗传效应水平间对比的不可估性。这两种定义似乎有些不连贯, 因为从理论上讲, 随机效应中不关联是不会出现的^[1]。

当进行场间或地区间的遗传评估时, 会遇到两个问题: 一是数量阶和位于比较元素间遗传联系不够的数量 (number), 这些因素可以导致不准确的估计; 二是性质上的和发生收集数据时受非控制因子的影响, 而产生有偏估计。

从数量的观点出发, 度量关联性应用 Laloë 等^[2]提出的决定系数法 (CD, coefficient of determination), 追求的精度由适当地定义对比向量 X 测量。为了检定连接数据的有效性, 当所有数据含一个或一些省略的关联组分时, 比较遗传估计值或场 (动物) 的秩是可能的。重要的变化将表明连接数据提供有偏估计。省略数据时可以用 Foulley 的方法进行度量^[1], 简化动物模型被简化数据取代。本文将采用这两种方法, 度量公畜模型下猪群体间的关联性。

1 材料与方法

1.1 数据结构

本研究收集了加拿大安大略省的 4 个猪种 (约

的克夏、长白猪、杜洛克和汉普夏)1987~1996年的52个猪场10个胎次的繁殖性能记录,各品种记录数和公猪数分别为:约克夏25478个和1456头、长白猪19424个和1026头、杜洛克3698个和428头和汉普夏2478个和253头。

1.2 数学模型

建立的混合模型为:

$$y = Xb + Zu + e \quad (1)$$

其中: y 为观察值向量; b 为固定效应向量; u 为随机效应向量; e 为随机残差向量; X 和 Z 分别为 b 和 u 的结构矩阵。

$$\begin{pmatrix} u \\ e \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} A \sigma_a^2 & 0 \\ 0 & I \sigma_e^2 \end{pmatrix} \right]$$

A 为亲缘系数矩阵, σ_a^2 和 σ_e^2 分别为加性和剩余方差组分。 u 的BLUP估计值表示为 \hat{u} ,而 $\hat{u} = (Z'MZ + A^{-1}\lambda)^{-1} Z'My$,其中: $\lambda = \sigma_e^2 / \sigma_a^2$ 和射影矩阵(projection matrix) $M = I - X(X'X)^{-1}X'$, $MX = 0$ 。

u 和 \hat{u} 的联合分布是多变量正态和零期望,方差矩阵等于

$$\Sigma_1 = \begin{pmatrix} A \sigma_a^2 & \Psi \\ \Psi & \Psi \end{pmatrix}$$

而 $\Psi = A \sigma_a^2 - (Z'MZ + A^{-1}\lambda)^{-1} \sigma_e^2 = A \sigma_a^2 - C^{uu} \sigma_e^2$ 。

$u | \hat{u}$ 和 $u - \hat{u}$ 属于多变量正态分布,分别表示为 $N(\hat{u}, C^{uu} \sigma_e^2)$ 和 $N(0, C^{uu} \sigma_e^2)$ 。

另有随机模型:

$$y = 1m + Zu + e \quad (2)$$

在这个随机模型中, $u | \hat{u} \sim N(\hat{u}, C_r^{uu} \sigma_e^2)$ 和 $u - \hat{u} \sim N(0, C_r^{uu} \sigma_e^2)$,其中 $C_r^{uu} = (Z'M_r Z + A^{-1}\lambda)^{-1}$ 和 $M_r = I - 1(1'1)^{-1}1'$,该射影矩阵正交向量等于1。为了预测育种值,该模型能被考虑显示数据提供的信息,由于固定效应的估计,除平均数外没有任何损失。

1.3 评定标准

两种方法或标准(criteria)用来评定一个对比预测的量,即育种值 $x'u$ 的线性组合,其中 x 是一个向量,元素的总和为0。

1.3.1 关联指数

关联指数(connectedness index)用 $IC(x)$ 表示^[1],即准确了解固定效应或不存在时(简化模型),预测误差方差(PEV, prediction error variance)^[3]相对减少。 IC 在0~1范围内变化,如果动物联系紧密,该值接近于1。其公式为:

$$IC(x) = x' C_r^{uu} x / x' C_r^{uu} x \quad (3)$$

1.3.2 广义决定系数

广义决定系数(generalized CD)用 $CD(x)$ 表示^[2],预测值与真值差异间相关系数的平方,反映遗传评估的精度,它等于:

$$CD(x) = 1 - [\text{var}(x'u | \hat{u}) \text{var}(x'u)] = x'(A - \lambda C^{uu}) x / x'Ax \quad (4)$$

1.4 总体指数

1.4.1 精度标准

注意遗传值间对比的 CD ,概括为最大特征值 μ_i :

$$[(A - \lambda C^{uu}) - \mu_i A] c_i = 0 \quad (5)$$

μ_i 在0和1之间: $\mu_2 \leq CD(x) \leq \mu_u$; μ_1 永远为0,特征向量 c_i 与 A^{-1} 成比例; $c_i A c_i = 0$,当 $i > 1 \Leftrightarrow 1'A^{-1} A c_i = 0 = 1'c_i$,即一个对比的定义; $CD(c_i) = \mu_{i0}$ 。

$x'u$ 可以表示为 c_i 的线性组合,等于

$$x'u = \left[\sum_{i=2}^n a_i c_i \right]' u \quad (6)$$

任何对比的 CD 是公式(5)特征值的加权均数:

$$CD(x) = \frac{\sum_{i=2}^n a_i^2 u_i}{\sum_{i=2}^n a_i^2} \quad (7)$$

两个总体精度指数(overall indices of precision)^[4] ρ_1 和 ρ_2 分别等于:

$$\rho_1 = \sum_{i=2}^n \frac{\mu_i}{n-1} \quad (8)$$

$$\rho_2 = \left[\prod_{i=2}^n \mu_i \right]^{\frac{1}{n-1}} \quad (9)$$

1.4.2 总体关联指数

由于 C^{uu} 是正值,所以关联指数不会为0。考虑一个对比 $x'u$,用模型(2)随机模型)获得 $x'u$ 的广义决定系数表示为 $CD_r(x)$ 。 $IC(x)$ 、 $CD(x)$ 和 $CD_r(x)$ 间关系可表示为:

$$IC(x) = \frac{1 - CD_r(x)}{1 - CD(x)} \quad (10)$$

这里, $x'u$ 的新指数用 $\phi(x)$ 表示,它等于 $CD(x)/CD_r(x)$ 。 $\phi(x)$ 在(不关联)与(不考虑固定效应)范围内变动。因此,总体关联指数(overall indices of connectedness)^[4] ϕ_1 和 ϕ_2 等于:

$$\phi_1 = \rho_1 / \rho_1 \quad (11)$$

$$\phi_2 = \rho_2 / \rho_2 \quad (12)$$

其中, ρ_{11} 和 ρ_{22} 是用随机模型(模型 2)求得的总体精度指数。

2 结果与讨论

采用公畜模型进行分析, 固定因子有场-年和胎次。产仔数的估计遗传力 (h^2) 为 0.108, $\lambda = (4 - h^2)/h^2 = 36.037$ 。经计算 4 个猪种的 ρ_{11} 、 ρ_{22} 和 ϕ_1 分别为: 约克夏猪 0.215、0 和 0.942, 长白猪 0.218、0 和 0.926, 杜洛克猪 0.135、0 和 0.878 以及汉普夏猪为 0.149、0 和 0.869, 表明加拿大安大略省的种猪群具有良好的遗传联系。在这 4 个猪种中, 只有少量公猪没有遗传联系, 即该公猪只在一个猪场使用, 其比例为: 约克夏猪 0.62% (9/1456), 长白猪 0.49% (5/1026), 杜洛克猪 2.80% (12/428) 和汉普夏猪 1.58% (4/253)。

根据被评估群体间或管理组 (management group) 遗传水平间不同的先验知识, 关联性的问题给予不同的定义。首先, 如果差异是/或假定是已知的, 它们能否可以展现在评估中? CD 能够回答此问题。其次, 不存在差异, 不关联只是降低精度的一个原因,

研究它没有多大的意义。该研究可以给予一个增加精度的选择, 或者通过关联性的增加, 或者通过数据提供信息量的增加。IC 则不完全适合这种研究, 主要因为它不总是展示不关联, 以及因为它随着从数据中获得信息量而减小。大的 IC 值归结于良好的关联性或缺乏的信息。遗传评估必须是精确和差别对待的。CD 既组合了数据结构和信息量, 又考虑了 PEV 和遗传变异性, 是一个选择判断遗传评估精度或最优化相应设计的好方法^[4]。

参 考 文 献:

- [1] Foulley J L, Hanocq E, Boichard D. A criterion for measuring the degree of connectedness in linear models of genetic evaluation[J]. Genet Sel Evol, 1992, 24: 315 ~ 330.
- [2] Laloë D. Precision and information in linear models of genetic evaluation[J]. Genet Sel Evol, 1993, 25: 557 ~ 576.
- [3] Kennedy B W, Trus D. Considerations on genetic connectedness between management units under an animal model[J]. J Anim Sci, 1993, 71: 2341 ~ 2352.
- [4] Laloë D, Phocas F, Ménessier F. Considerations on measures of precision and connectedness in mixed linear models of genetic evaluation [J]. Genet Sel Evol, 1996, 28: 359 ~ 378.