

- ness in linear models of genetic evaluation [J]. *Genet Sel Evol* 1992, 24: 315–330.
- [8] ROSO V M, SCHENKEL F S, MILLER S P. Degree of connectedness among groups of centrally tested beef bulls [J]. *Can J Anim Sci* 2004, 34: 37–47.
- [9] FRIES L A. Connectedness in beef cattle genetic evaluation: The heuristic approach used in MLC FOR [A]. *Proc 6th World Congr Genet Appl Livest Prod*, Armidale NSW, Australia Vol 27 [C]. 1998: 449–500.
- [10] MATHUR P K, SULLIVAN B R, CHESNAIS J P. Measuring connectedness: Concept and application to a large industry breeding program [A]. *Proc 7th World Congress on Genetics Applied to Livestock Production Montpellier France Communication No. 20–13 [C]*. 2002.
- [11] MATHUR P K, SULLIVAN B R, CHESNAIS J P. Estimation of the degree of connectedness between herds or management groups in the Canadian swine population [EB/OL]. Canadian Centre for Swine Improvement Ottawa Canada (Ministry). 1998.

(Edited by LI XIAO HUI)

译文

场间“关联”在场间遗传评估中的作用

P. K. MATHUR

(Canadian Center for Swine Improvement Central Experimental Farm
Building 54 Maple Drive Ottawa ON K1A 0C6 Canada)

摘要: “关联”是遗传评估的重要组成部分, 关联度影响各场个体或各同期组的估计育种值 (EBVs) 比较的准确性。可以用关联率 (CR) 来度量, 它是利用同期组效应估计值的方差和协方差来计算的, 目前已有相应的算法和计算程序。CR 的最低需要量依赖于同期组的大小、准确性要求和残差方差。如果对于同期组约有 100 个个体, 要检测出的 EBVs 差异大于性状标准差的 20%, 则 CR 要达到大约 48%。如果同期组较小, 要求更加准确的比较, 则需要更高的 CR。因此, 对于参加同一个遗传评估方案的育种者来说, 要进行优秀遗传资源的交换, 有可能的话, 使用一些共同的测定设备, 以使育种值的估计结果有意义。维持良好的关联水平, 可以使选择优秀种猪的遗传评估方案更有价值, 获得更快的遗传进展。

关键词: 关联; 育种值; 遗传评估

1 遗传关联的概念

在动物育种中, “关联”的概念主要应用于遗传评估和群间 EBV 的比较。在这里, 关联可以用来度量群间评估的精确性。在比较不同场或不同群体间个体遗传值估计准确性的影响, “关联”可定义为场间或同期猪群间的相关。

从统计学的角度来看, 如果模型里主效应中的每个对应的相同固定效应都可以估计的话, 则该数据集是关联的^[1]。在遗传评估模型, 尤其是 BIUP 模型中, 场效应或同期组效应构成一对固定效应, 如果要比较这些组间个体的差异, 则场效应或同期组效

应必须是可估的。次级组间个体差异估计的准确性受到次级组间关联度的影响。

育种值估计的准确性通常被当作“重复力”, 它依赖于性状的遗传力, 以及评估时可用的信息量, 比如个体和亲属记录的数量, 它反映了用估计育种值 (EBV) 对真实育种值进行估计的可靠性, 但是它并不能反映不同期组 EBV 比较的精确性, 这一准确性依赖于同期组间的关联度。

关联的概念可用于同一场或不同场的同期组个体育种值的比较。在制订共同的遗传评估计划时, 经常需要知道不同场个体之间的关联是否足够。它对不同场的个体育种值的排队, 以及在选择和遗传改

良中挑选优秀个体都非常有用。这时，每一个场的个体可以看成是同一个同期组。如果一个场有数个同期组，可以选择某一个同期组代表该场。比如，一个场有数个场-年-季组别，最近的场-年-季组可以代表该场。下面为了简便起见，我们用场来表示这些同期组。

1.1 关联的统计学标准

在遗传评估中应用的关联概念如图1所示。场A和场B有关联，因为它们都使用了种猪1；场B和场C有关联，因为它们都使用了种猪2；但场C和场D是没有关联的，因为它们之间没有共同的遗传联系（genetic link）。因此，场B和场C的个体育种值差异可以估计和比较，而场C和场D的个体育种值差异的比较则准确性不够，因为它们之间没有关联。事实上，场间个体的育种值进行比较时，没有必要一定要有直接的关联，场间个体之间也可以有间接的遗传联系。在本例中，场A和场C因为有种猪1和2而存在间接的联系。即使不同场的个体之间没有遗传关联，场间的比较有时也是可以的。比如，场C和场D的个体在同一个测定站内进行测定，则场C和场D也存在关联。这样，场A、B、C和D就关联起来了。关联是一种统计概念，有关联则效应可比，认识到这一点很重要。Kennedy和Tus[2]，Hanocq等[3]也提出了同样的观点。遗传联系不一定非要有关联存在，不过，遗传联系很重要，也是提高关联度的有效途径。

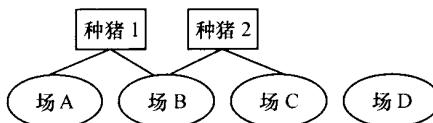


图1 通过遗传联系建立场间关联

1.2 关联度

在遗传评估的模型中，可以通过将场效应作为随机效应而不是固定效应，用于场间EBV差异的估计。然而即使不同场个体EBV的差异是可估的，存在的主要问题还是比较的准确性，尽管把这种场间效应当作随机效应并且它们是可估的，场间的关联度仍然影响这种比较的准确性。因此，知道关联度就比只知道场间是否有关联更重要而且有用。场间可能存在关联，而且场间个体的EBV可比，但如果场间关联较低，则这种比较的准确性也不高，可能就与完全没有关联时的准确性相似。事实上，在利用动物模型BLUP进行遗传评估时，通过分子亲缘相关矩阵、遗传分组和其他固定效应和随机效应，就会产生

群间关联^[4,5]。但是，应该认识到关联度是影响场间个体比较准确性的重要因素。

2 关联的度量

2.1 EBV差异的预测误差方差

估计方法依赖于估计的应用目的。动物育种中的主要目的是知道不同场或同期组个体EBV差异比较的准确性。因此，理想的度量方法应该依据场内个体EBV或不同场个体EBV差异比较的准确性。个体EBV的准确性通常用预测误差方差来度量，后者为混合模型方程组中个体效应对应的逆矩阵中的元素。2个个体EBV差异比较的准确性可以用该差异的预测误差方差（PEVD）来表示：

$$\text{PEVD} = \text{Var}(\hat{a}_i - \hat{a}_j) = \text{Var}(\hat{a}_i) - 2\text{cov}(\hat{a}_{ij}).$$

来自不同场2个个体EBV差异的平均准确性可以用这2个场所有个体两两配对比较的平均PEV来表示。可是，在大规模的遗传评估计划中，可能有一个场和另一个场的数个个体配对，也可能有数个场的配对比较，这些计算非常耗时。因此，需要一种较简单的方法。

2.2 关联率（CR）

一个有本身记录的个体EBV取决于由同期组所估计的生产性能差异、场效应和遗传评估模型中的其他因子。因此，不同场或不同组的2个个体EBV差异的PEV可以分成2部分：一部分与个体本身及其亲属有关，这部分取决于每个EBV的PEV，2个个体的相关程度、个体与场效应之间的协方差；另一部分与场效应本身有关，包括场效应估计值的方差和不同场效应间的协方差。计算2个场个体的PEVD时，对于每个配对来说，第2部分中与场效应有关的部分是相同的；而与个体本身及其亲属有关的部分，不同配对是不同的，而且所有的配对进行平均后会相互抵消。如下式所示：

$$\begin{aligned} \text{VED}_{ij} &= \text{Average}[\text{PEV}(\hat{a}_{ik} - \hat{a}_{jk})] \approx \\ &\text{Var}(\hat{h}_i - \hat{h}_j) \triangleq \text{Var}(\hat{h}_i) + \\ &\text{Var}(\hat{h}_j) - 2\text{cov}(\hat{h}_{ij}), \end{aligned}$$

结果，所有比较的PEV平均值近似等于2个场效应估计值的PEV。因此，场效应差异的PEV可以用来度量场间关联的高低。Kennedy和Tus^[2]的模拟研究表明，它可以用来度量关联程度的高低，PEVD和VED之间的相关程度很高（0.995）。这些场间效应方差和协方差可以用下面的公式来计算关联率（CR），用于度量关联的高低：

$$\text{CR}_{ij} = \frac{\text{cov}(\hat{h}_{ij})}{\sqrt{\text{Var}(\hat{h}_i)\text{Var}(\hat{h}_j)}},$$

上式中, $Var(h_i)$, $Var(h_j)$ 和 $cov(h_i, h_j)$ 分别是相应场效应的方差和协方差。这样, CR似乎与场效应的相关相似, 不过, 它却与简单的积矩相关有很大差异。

计算 CR时, 也需要计算场效应的方差和协方差。通常很难直接由 MME 的逆矩阵中得到与场效应对应的逆矩阵元素。因此, 用下面的方法, 根据矩阵理论, 矩阵与其逆的积为单位矩阵, 这点对于 MME 的系数矩阵 $(w'w)$ 也是正确的。

$$w'w (w'w)^{-1} = I$$

因此

$$w'w (w'w)_i^{-1} = I_i$$

这里, $w'w = \text{MME}$ 的系数矩阵; $I = \text{单位矩阵}$; $I_i = \text{对应于同期组 } i \text{ 的单位矩阵向量}$ (在该组对应位置为 1 其他为 0); $(w'w)_i^{-1} = \text{与第 } i \text{ 个同期组对应的逆矩阵元素向量}$ 。

如果用迭代法求解 MME, 可以从相应的向量 I_i 得到所需的场效应的某行或某列的逆矩阵元素。这些向量可以用包含相应场效应方差和协方差的逆元素组成的分块矩阵表示。事实上, 可以用更一般的方法来计算特定的逆元素, 而无需直接计算出 MME 的所有元素的逆。

在过去 7 年里, 将这一 CR 计算方法和估计结果用于参加加拿大猪改良计划 (CCSI) 的大约 100 个猪群, 结果表明 CR 用于表示关联水平的高低很理想, 利用它可以指导育种者提高 CR, 使之达到比较理想的程度。

2.3 估计关联的其他方法

关联还有其他估计方法, 一种是将关联作为一个统计量, 另外一种是只用于度量遗传联系。Labé 等^[6] 用估计 VED 比较了 2 种计算关联的方法, 它们分别是: 真实遗传值差异与预测值间相关系数的平方 (CD), 以及 Foulley 等^[7] 提出的反映固定效应已知时 PEV 相对下降的关联指数 (IC)。结果表明, CD 结合了 VED 和遗传变异性, 这样, CD 既反映了关联, 又反映了遗传变异性。该方法可以用来控制遗传变异, 尤其是育种群较小有可能导致出现近交系数问题等的情况下。可是, 如果目的是得到关联的高低, 以及提供有关场的信息以决定是否需要提高关联水平, 则计算出 CR 以估计关联度可能更有用。因为人工授精 (AI) 的广泛应用, 可以在国际间利用优秀种畜, 从而导致了遗传变异的降低。这是过去奶牛育种存在的主要问题, 目前在猪育种同样也遇到了这个问题。关联可以用来提供有关每头公畜有效后

裔、有效群体含量、近交水平等方面的信息。

在一个肉牛测定群的研究中, Roso 等^[8] 比较了 4 种估计关联度的方法, 这些方法包括估计育种值差异的预测误差方差 (PEVD)、测定组效应差异的估计方差 (VED)、有直接遗传关联的个体总数 (GLT)^[9]、关联率 (CR)。结果表明, 某测定组与其他所有测定组的平均 PEVD 可以根据 CR 或 GLT 得到较好地估计。在用于肉牛测定组的数据集里, 与 VED 和 CR 相比, 只根据遗传关联个体数计算出来的 GLT 也可以提供关联的估计值。

3 CR 的计算程序

在 CCSI 的网页 (www.ccsi.ca/connectedness) 上有计算 CR 的程序。该程序是根据动物育种工具箱 (ABTK) 编写的。ABTK 是主要用 BLUP 方法进行遗传评估的一组程序, 其中的一些小程序可以用来完成各种简单功能, 如矩阵求逆、矩阵相乘和建立关联矩阵等。这些程序最早是用 UNIX 和 Linux 操作系统开发的。现在 ABTK 也可以在 Windows 操作系统下运行^[10]。在应用这个程序计算 CR 时, 需要输入系数矩阵文件、包含场效应的文件、含有场标识的文件等。该程序输出的是 MME 一个块的逆和每个配对场效应间的 CR。

4 场间遗传比较时需要的关联水平

人们经常要问这样一个问题, 即进行有效的场间比较时需要的关联水平要有多高。利用最简单的方法可以回答这个问题, 需要的关联高低取决于准确性的要求, 这可以通过下面的 t 检验来决定:

$$t_a < \frac{d}{\sqrt{VED(h_i - h_j)}},$$

式中, t_a = 理想显著水平时 t 统计量的值; d = 需要检出的 EBV 平均差异值; $VED(h_i - h_j)$ = 场间差异估计值的方差。

根据 Mathur 等^[11], CR 和 VED 的相关程度可由下式计算:

$$\begin{aligned} VED(h_i - h_j) &= Var(h_i) + Var(h_j) - \\ 2 cov(h_i, h_j) &= Var(h_i) + Var(h_j) - \\ &\quad 2 CR_{ij} \sqrt{(Var(h_i)) (Var(h_j))}, \end{aligned}$$

因此, 需要的 CR 可以由下式计算:

$$CR_{ij} > \frac{[Var(h_i) + Var(h_j)] - \frac{d^2}{t_a^2}}{2 \sqrt{Var(h_i) Var(h_j)}},$$

因此, 根据管理组的大小, $VED(h_i - h_j)$ 可近似地写为:

$$VED(h_i - h_j) \cong \sigma_e^2 \left(\frac{1}{n_i} + \frac{1}{n_j} - 2 \frac{CR_{ij}}{\sqrt{n_i n_j}} \right),$$

这里, σ_e^2 =性状的残差方差; n_i 和 n_j 分别是对应的管理组或场的个体数量.

根据这种定义, 在显著水平为 α 的情况下达到显著差异 d 的 t 统计量为:

$$t_\alpha < \frac{d}{\sqrt{\sigma_e^2 \left[\frac{1}{n_i} + \frac{1}{n_j} - 2 \frac{CR_{ij}}{\sqrt{n_i n_j}} \right]}},$$

需要的关联水平可以通过下式获得:

$$CR_{ij} > \sqrt{n_i n_j} \frac{1}{2} \left[\left(\frac{1}{n_i} + \frac{1}{n_j} \right) - \frac{d^2}{\sigma_e^2 (t_\alpha)^2} \right],$$

由上式可知, 需要的关联水平取决于场或同期组的大小 (n_i 和 n_j)、显著水平 (α)、需要检测出的 EBV 平均差异值 (d)、以及残差方差 σ_e^2 .

如果 2 个场的大小相同 ($n_i = n_j = n$):

$$CR_{ij} > 1 - \frac{n}{t_\alpha^2} \left(\frac{d}{\sigma_e} \right)^2,$$

这时 CR 的需要量是样本大小、显著水平、需要检测出的差异高低与残差标准差的函数.

考虑猪育种中一种典型的情况, 假定选择指数以标准化, 其平均数等于 100 标准差等于 25. 如果需要检测出指数差异超过 5 的差异, 即为指数标准差的 20%. 进一步假定需要的显著水准是 $P < 0.05$. 2 个猪群的大小等于 100, 则关联要大于或等于 48%. 这种情况下, 需要的 CR 高低取决于样本大小 ($n_i = n_j = n$)、需要检测出的差异大小与残差标准差之比. 在 $P < 0.05$ 时的结果如图 2 所示.

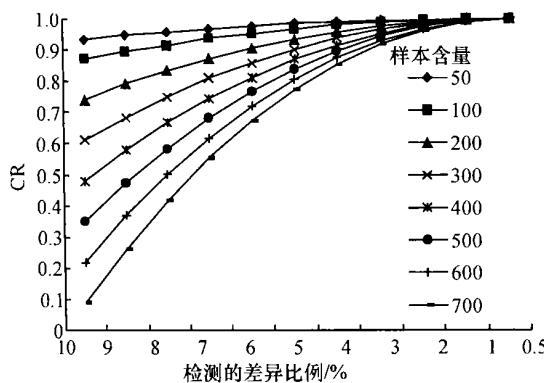


图 2 不同样本含量和准确性要求时关联率 (CR) 的需要量

需要检测出的差异与性状标准差之比越小, 则需要的 CR 越高. 在样本含量很小时尤其如此. 另外一种确定 CR 的可行途径是检查不同配对 CR 的差异对 EBV 差异标准误的影响. 图 3 所示的是, 在大群体中, 不同配对 CR 与不同场效应差异的标准误 (VED 的平方根) 的关系.

图 3 中显示出这 2 个变量间为非线性关系. 随

着 CR 的上升, 差异标准误下降. 当许多配对比较的 CR 都很低, 特别是 CR 低于 5% 时, 误差尤其高. 因此, 这个数字可以看成是场间合理比较时关联率的最低需要量.

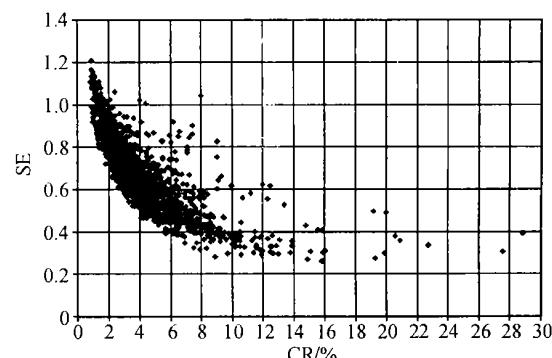


图 3 产仔数性状的差异标准误 (SE) 和关联率 (CR) 的关系

关联是在统计学上用来度量场效应估计结果的准确性的, 因此, 与群体大小有关. 当群体大小低于 50 时, 群体大小对 CR 的影响较大^[10]. 随着样本含量的上升, CR 的大小越来越依赖于遗传联系, 而样本含量的影响变小.

5 如何增加关联

最简单的提高关联的途径是在不同的场和同期组利用共同的种猪. 在关联程度很好的畜群中应用有后裔的 AI 公猪可以提高关联. 提高关联时考虑下面的因素很重要: 后裔的数量、这些后裔所分布的群体数量、几个场的加权关联率等^[11]. 通常大约 15% 的后裔个体来自关联度较高的公猪就足以达到关联水平的最低要求. 没有后裔的年轻公猪也可以用于 AI, 只要它们来自关联较高的群体.

利用中心测定站对来自不同场猪, 在同一个环境和管理条件下进行测定, 也是提高关联的一种方法, 尤其是当遗传交换很少的情况下.

估计了 EBV 的同期组足够大也很重要. 如果某个同期组同一个品种的个体低于 10, 则 EBV 的准确性也很低. 在实践上, 每个同期组或场一年一季组至少要有 10 个个体. 这 10 个以上的个体应该至少是来自 3 头以上的公猪后代, 以达到足够的场内关联, 以及估计准确性的要求.

为了持续保持遗传性能良好的公猪有密切的关联, 要求参加育种计划的场利用本场优秀的种猪, 或组成优秀 AI 公猪群. 这种类型的育种体系使国家育种计划类似于一个大的核心群, 具有不同场的遗传变异和良好的关联. 这种育种结构可以进行群间比较, 提供高的选择差, 获得大的遗传进展.

参考文献 (见原文)