第 27 卷 第 4 期 作 物 学 报 Vol 27, No. 4
2001 年 7 月 ACTA AGRONOMICA SINICA July, 2001

品种区域试验中算术平均值、BLUP 和 AMMI 估值的精度比较*

张群远[1] **孔繁玲**[1] **杨付新**[2]

([1]中国农业大学植物遗传育种系, 北京 100094; [2]中国农业科学院棉花研究所, 河南安阳 455112)

提 要 利用 1982 年以来我国棉花、小麦、水稻和玉米的 60 套区域试验数据, 采用交叉验证方法, 对区域试验中算术平均值、最佳线性无偏预测值(best linear unbiased predictor, BLUP)和AMMI(additive main effects and multiplicative interaction)模型估值的预测精度进行比较, 结果表明, 与算术平均值相比, AMMI 估值精度的增益倍数(gain factor, GF)平均为 1.045, 变幅为 0.963～1.414, 其精度多数情况下提高不大; BLUP 的 GF 平均为 1.170, 变幅为 1.008～1.619, 其精度普遍较高。同时, 文中对 3 种估值的模型作了论述和比较。

关键词 区域试验; BLUP; AMMI; 预测精度

Comparison of the Predictive Accuracy of Arithmetic Means and BLUPs and AMMI Estimates in Regional Crop Trials

ZHANG Qun-Yuan[1] KONG Fan-Ling[1] YANG Fu-Xin[2]

([1] *Department of Plant Genetics and Breeding, China Agricultural University, Beijing* 100094; [2] *Institute of Cotton, Chinese Academy of Agricultural Science, Anyang* 455112, *China*)

Abstract Sixty sets of data from regional trials of cotton, wheat, rice and maize since 1982 in China were used in cross validation to compare the predictive accuracy of arithmetic means and BLUPs (*best linear unbiased predictors*) and AMMI (*additive main effects and multiplicative interaction*) estimates. The average precision gain factor (GF) of AMMI relative to arithmetic mean was 1.045 with a range from 0.963 to 1.414, which showed slight increases of precision; BLUP was found commonly to outperform arithmetic mean and AMMI with an average GF of 1.17, ranged from 1.008 to 1.619.

Key words Regional Trial; BLUP; AMMI; Predictive Accuracy

作物品种区域试验的目的是在多环境下对参试品种进行比较和评价, 以确定新品种的推广价值和适应范围。这种比较和评价有赖于对各品种在各环境下的产量(或其它性状值, 统称品种×环境组合均值)作出准确估计。我国区试中历来采用算术平均值进行估计。算术平均值是最为常用的一种估值, 虽简便易行, 但存在一定局限。一方面, 由于算术平均值是直接以样本平均数来估计总体均值, 估计时未能充分利用试验中多种变异的信息, 要获得准确

* 国家自然科学基金资助项目(30070433)。赵虹、王磊、王洁、葛知男和孙世贤同志提供了部分数据, 特此致谢!
收稿日期: 2000-10-08; 接受日期: 2000-12-24
Received on: 2000-10-08; Accepted on: 2000-12-24

估值, 需较多的重复数; 另一方面, 算术平均值实质上是对品种试验表现的事后描述, 并非对品种未来表现的预测, 而后者是我们真正感兴趣的, 具有更重要的实践意义。近年来, 国外不少研究表明, 利用一些较复杂的模型和方法, 可以得到预测精度比算术平均值更高的估值, 其中混合线性模型(mixed linear model)的最佳线性无偏预测(best linear unbiased prediction, BLUP)[1]和AMMI(additive main effects and multiplicative interaction)模型的AMMI估值[2]是最主要的两种。本文结合我国多年区域试验的数据, 对算术平均值、BLUP 和AMMI三种估值的模型和预测精度进行比较, 探讨各种方法在我国区试中的适用性, 以期有针对性地引入和利用。

1　材料和方法

1.1　数据资料

数据取自我国棉花、小麦、水稻和玉米区试分重复记载的共 60 年次的历史资料(详见表1)。其中棉花数据为皮棉产量, 小麦、水稻和玉米数据为子粒产量, 单位均为 kg/hm²。

1.2　各种估值的模型和计算

表1　区试数据及其提供单位

Table 1　Data and their providers

作物 Crops	区试组别 Trial groups	年份 Years	资料提供单位 providers of data
棉花 Cotton	黄河流域春棉	82～96	中国农科院棉花所
	黄河流域夏棉	86～93	中国农科院棉花所
	长江流域常规棉	90～98	江苏农科院经作所
小麦 Wheat	黄淮春水组 黄淮冬水组	87～91	河南农科院小麦所
水稻 Rice	中早粳晚熟组	92, 93, 96, 97, 98	中国农科院作物所
	中早粳中熟组	93, 94, 98	中国农科院作物所
	中籼迟熟组	91, 94, 95, 96	中国农科院水稻所
玉米 Maize	华北春玉米	96～98	全国农技推广服务中心
	西北春玉米	96, 97	
	黄淮夏玉米	96	

为便于论述, 下面以 v 个品种, s 个环境(通常是地点或地点×年份的组合环境), r 次重复的区域试验为例来阐明各种估值所依据的数学模型及具体计算方法。第 i 个品种在第 j 个环境中的第 k 次重复观测值记为 Y_{ijk}。各种估值的计算就是依据特定模型对 Y_{ijk} 中的信息进行处理和提取的过程; 而各种模型的实质是根据各种假设把 Y_{ijk} 表达成相应的理论构成。

1.2.1　算术平均值　　用算术平均值来估计各品种在各环境下的性状值, 所依据的是统计学上最简单的一种线性模型:

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \tag{1}$$

μ_{ij} 为第 i 个品种在第 j 个环境中的真实均值, 也就是我们要估计的真值; ϵ_{ijk} 为第 i 个品种在第 j 个环境中的第 k 次重复观测值的误差。从试验设计的角度看, 区试中品种和环境是两个因素, 品种 i 和环境 j 的搭配即为一个处理, μ_{ij} 也就是处理均值。所以, 此模型也叫处理均值(treatment means)模型[2]; 它把每个 Y_{ijk} 表示为 μ_{ij} 加上 ϵ_{ijk} 的形式, 并用算术平均的方法对 μ_{ij} 作出估计:

$$\hat{\mu}_{ij} = \overline{Y_{ij}} = \sum_{k=1}^{r} Y_{ijk}/r \tag{2}$$

$\overline{Y_{ij}}$ 即为第 i 个品种在第 j 个环境中的算术平均值。这一模型十分简单和直观, 实际应用中往往不对其作专门说明。

1.2.2　最佳线性无偏预测值(BLUP)　　区试中由于方差分析的需要, 更常用的是另一种线性可加模型[2]:

$$Y_{ijk} = \mu + g_i + e_j + \theta_{ij} + \epsilon_{ijk} \tag{3}$$

μ 为所有观测值所属总体的均值; g_i 为品种 i 的效应; e_j 为环境 j 的效应; θ_{ij} 为品种 i 与环境 j 的基因型×环境(GE)互作效应; ϵ_{ijk} 同(1)式。区试中往往根据这一模型来进行方差分析。这一模型实质上是把(1)式模型中的 μ_{ij} 进一步分解为 μ、g_i、e_j 和 θ_{ij} 四种构成, 此时, μ_{ij} 为四种效应的一个可估函数(即四者之和); 利用更一般的线性模型求解方法, 可以获得 μ_{ij} 的最佳线性无偏估值(best linear unbiased estimation, BLUE)或最佳线性无偏预测值(BLUP)[3]。在模型求解之前, 首先要对各种效应是随机还是固定作出假设。“固定”意味着该效应值在试验中是一系列定值; “随机”则是指试验中该效应的一系列值是来自于具有特定均值和方差的总体的一个随机样本。一般来说, μ 是固定效应, ϵ_{ijk} 是随机效应, g_i、e_j 和 θ_{ij} 则可根据实际情况作出各种假设, 所以, (3)式往往是一个既有固定效应, 又有随机效应的混合线性模型。在 μ_{ij} 的四种构成均为固定效应时, 其估值为 BLUE; 在四种构成中含有随机效应时, 由于 μ_{ij} 不再是一个严格意义上的参数, 故其估值称为预测值, 即 BLUP。Peipho(1994)[1]曾在 Henderson(1975)[3]基础上推导出平衡数据时各种模型下 μ_{ij} 的 BLUE 或 BLUP 的计算公式如下:

模型Ⅰ: g_i、e_j 和 θ_{ij} 均为固定效应; μ_{ij} 的估值记为 BLUE;

$$\mathrm{BLUE} = \overline{Y} + (\overline{Y}_i - \overline{Y}) + (\overline{Y}_j - \overline{Y}) + (\overline{Y}_{ij} - \overline{Y}_i - \overline{Y}_j + \overline{Y}) = \overline{Y}_{ij} \tag{4a}$$

模型Ⅱ: g_i、e_j 和 θ_{ij} 均为随机效应; μ_{ij} 的估值记为 BLUPge;

$$\mathrm{BLUPge} = \overline{Y} + h_g(\overline{Y}_i - \overline{Y}) + h_e(\overline{Y}_j - \overline{Y}) + h_{ge}(\overline{Y}_{ij} - \overline{Y}_i - \overline{Y}_j + \overline{Y}) \tag{4b}$$

模型Ⅲ: e_j 为固定效应, g_i 和 θ_{ij} 为随机效应; μ_{ij} 的估值记为 BLUPg。

$$\mathrm{BLUPg} = \overline{Y} + h_g(\overline{Y}_i - \overline{Y}) + (\overline{Y}_j - \overline{Y}) + h_{ge}(\overline{Y}_{ij} - \overline{Y}_i - \overline{Y}_j + \overline{Y}) \tag{4c}$$

模型Ⅳ: g_i 为固定效应, e_j 和 θ_{ij} 为随机效应; μ_{ij} 的估值记为 BLUPe。

$$\mathrm{BLUPe} = \overline{Y} + (\overline{Y}_i - \overline{Y}) + h_e(\overline{Y}_j - \overline{Y}) + h_{ge}(\overline{Y}_{ij} - \overline{Y}_i - \overline{Y}_j + \overline{Y}) \tag{4d}$$

其中:

$$h_g = \frac{\sigma_{GE}^2 + s\sigma_G^2}{\sigma^2/r + \sigma_{GE}^2 + s\sigma_G^2} \tag{5a}$$

$$h_e = \frac{\sigma_{GE}^2 + v\sigma_E^2}{\sigma^2/r + \sigma_{GE}^2 + v\sigma_E^2} \tag{5b}$$

$$h_{ge} = \frac{\sigma_{GE}^2}{\sigma^2/r + \sigma_{GE}^2} \tag{5c}$$

以上各式中, $\overline{Y}$ 为试验总均值; $\overline{Y}_i$ 为第 i 个品种的试验均值; $\overline{Y}_j$ 为第 j 个环境的试验均值; $\overline{Y}_{ij}$ 同(2)式; σ^2、σ_G^2、σ_E^2 和 σ_{GE}^2 分别为误差、品种、环境以及品种×环境互作的方差; 均衡数据时, 可以按一定模型假设进行方差分析, 通过求解期望均方组成来估计这些方差值[4]。用这些方差估值代替真值后得出的结果, 虽然已不再是严格的 BLUP, 但习惯上仍称之为 BLUP。

根据(4a)式可看出, 算术平均值其实就是固定模型下 μ_{ij} 的 BLUE。由(4b)、(4c)和(4d)式不难看出, BLUP 实质是依据随机效应方差和误差方差的大小, 相应减小了随机效应在 μ_{ij} 估值中所占的比例, 对算术平均值作了适当的“收缩”。模型Ⅱ和Ⅲ中把品种效应 g_i 看作随机的, 这似乎与我们区试中方差分析时通常采取的品种效应固定的习惯做法有所矛盾。事实上, 只要品种效应值服从一定的概率总体的分布, 即使试验方案中品种并非随机抽取, 依据特定的分析目的, 在统计上也可作为随机效应看待[1]。就实际含义来看, (5a)式意味着, 若试验误差 σ^2 越大, 对试验中表现越极端的品种(即 $\overline{Y}_i - \overline{Y}$ 的绝对值越大), 越应该持“谨慎”态度

(即 $\overline{Y_i}-\overline{Y}$ 的"缩小"越多)，这与我们的实际经验是吻合的。所以，品种随机的假设在此有其合理性。

1.2.3　AMMI 模型及其估值　对模型(3)中的互作效应值 θ_{ij} 进行主成分分解(principal component analysis, PCA)，即得到 AMMI 模型[2]:

$$Y_{ijk} = \mu + g_i + e_j + \sum_{n=1}^{p}\lambda_n u_{in} v_{jn} + \rho_{ij} + \epsilon_{ijk} \tag{6}$$

λ_n 为品种×环境两向互作值矩阵的第 n 个奇异值(其平方即为特征根); u_{in} 为互作值矩阵中品种 i 的特征向量的第 n 个元素; v_{jn} 为互作值矩阵中环境 j 的特征向量的第 n 个元素; ρ_{ij} 为 θ_{ij} 进行 PCA 分解后的剩余部分; p 的最大可取值 N 为 s 和 v 中的最小值减 1，即 $p \le N = \mathrm{Min}(s-1, v-1)$。p 的取值不同，(6)式可得到不同的模型，所以，AMMI 模型其实是一系列模型的总称。p 取值从到 N，对应模型分别称为 AMMI-0、AMMI-1、AMMI-2……AMMI-N 模型。实际应用中往往根据 F 测验的显著性来确定 p 的大小(一般取 λ_n 值较大的前 1～3 项)，这时，p 以后的(N-p)个 PCA 项被当作剩余归入 ρ_{ij} 中。AMMI 模型中的 ρ_{ij} 被当作误差看待，所以，品种×环境组合均值 μ_{ij} 的 AMMI 组成为:

$$\mu_{ij} = \mu + g_i + e_j + \sum_{n=1}^{p}\lambda_n u_{in} v_{jn} \tag{7}$$

由于 AMMI 模型把各种效应看作是固定的，所以，与(4a)式同理，(7)式中的 μ、g_i 和 e_j 分别以 $\hat{\mu}=\overline{Y}$，$\hat{g}_i=(\overline{Y_i}-\overline{Y})$ 和 $\hat{e}_j=(\overline{Y_j}-\overline{Y})$ 进行估计; λ_n、u_{in} 和 v_{jn} 则需先求出互作值矩阵 $\hat{\theta}_{ij}=\overline{Y_{ij}}-\overline{Y_i}-\overline{Y_j}+\overline{Y}$，再作特征分解求得它们的估值。最后得到 μ_{ij} 的 AMMI 估值为:

$$\mathrm{AMMI}(\mu_{ij}) = \overline{Y_i} + \overline{Y_j} - \overline{Y} + \sum_{n=1}^{p}\hat{\lambda}_n \hat{u}_{in} \hat{v}_{jn} \tag{8}$$

AMMI 模型在保留模型(3)加性主效部分的同时，利用 PCA 方法压缩简化了互作的信息。理论上讲，这有助于剔除算术平均值中所包含的部分误差，从而提高 μ_{ij} 估值的精度。需要指出的是，当 AMMI 模型的 p 取最大值 N 时(此时称为 AMMI 全模型)，模型中乘式互作项总和与(3)式中的 θ_{ij} 一致，所以，此时 AMMI(μ_{ij}) 与(5a)式的 BLUE(μ_{ij}) 以及算术平均值相等。

1.3　各种估值精度的交叉验证

各种估值精度的比较采用交叉验证(cross validation)方法[2]。具体做法是，对每年的区试数据(v 个品种，s 个地点，r 次重复)，以试点为单位，把 r 个重复观测值随机分开，其中 r-1 个用于建立模型和估计 μ_{ij}(称为建模数据)，剩余 1 个用于验证(称为验证数据)。利用建模数据，根据(4)组式和(8)式，分别求出 μ_{ij} 的 BLUE、BLUPge、BLUPg、BLUPe 和 AMMI 估值(算术平均值即 BLUE，故不再单独计算); 由于本文使用的是平衡数据，(5)组式中的各种方差成分直接通过方差分析的方法获得[4]。对于每种估值，先根据(9)式出计算相应的平均预测差平方和(mean square prediction differences, MSPD)，然后根据(10)式求出该估值相对于算术平均值的精度增益倍数(gain factor, GF)。

$$\mathrm{MSPD} = \sum_{i=1}^{v}\sum_{j=1}^{s}(\hat{Y}_{ij} - Y'_{ij})^2 / vs \tag{9}$$

$$\mathrm{GF} = \frac{\mathrm{MSe}}{(r-1)(\mathrm{MSPD}-\mathrm{MSe})} \tag{10}$$

公式中 $\hat{Y}_{ij}$ 为品种 i 在环境 j 上的估计值; Y'_{ij} 为验证观测值; MSe 为环境内的误差项均方，由全部观测值的方差分析获得。重复进行 3000 次数据分样和计算，得到每种估值的平均

MSPD 和 GF。MSPD 反映了估计值和验证观测值之间的接近程度;GF 则意味着某估值的精度相当于算术平均值精度的倍数。所以,MSPD 越小,GF 越大,表明估值精度越高。上述过程中,采用QR 算法获得AMMI模型的奇异值和特征向量[5]。具体计算在微机上利用VB5.0编程实现。

2 结果与分析

针对表1中共60套一年多点的平衡数据,按照上述方法对BLUE、BLUPge、BLUPg、BLUPe和AMMI估值分别进行60轮交叉验证,并统计各种作物和区试组别的精度增益倍数(GF)的均值和变幅,列于表2。水稻和玉米由于数据较少,所以未分组别进行统计。另外,AMMI模型随p 取值的不同,可以得到多种AMMI估值,表中AMMI是指各轮验证中GF最大的一个,它代表AMMI系列模型在配合数据时所达到的最高精度。

表2 60次区试中的BLUE、BLUP和AMMI估值的交叉验证结果

Table 2 Cross validation results of BLUE, BLUP and AMMI in 60 trials

作物 Crops	区试组别 Trial groups	五种估值的精度增益倍数的平均值和变幅 Means and ranges of Precision gain factors(GF) of 5 kinds of estimators				
		BLUE	BLUPge	BLUPg	BLUPe	AMMI
棉花 Cotton	长江流域常规棉 Common cotton in Changjiang River region	1.001 0.984~1.018	1.208 1.102~1.424	1.204 1.101~1.407	1.207 1.102~1.420	1.054 0.989~1.273
	黄河流域春棉 Spring cotton in Yellow River region	1.003 0.972~1.033	1.221 1.024~1.402	1.216 1.022~1.394	1.220 1.024~1.400	1.025 0.972~1.081
	黄河流域夏棉 Summer cotton in Yellow River region	1.002 0.962~1.031	1.105 1.037~1.166	1.102 1.034~1.164	1.105 1.036~1.166	1.028 0.985~1.098
	棉花(总) Cotton(total)	1.002 0.962~1.033	1.188 1.024~1.424	1.184 1.022~1.407	1.188 1.024~1.420	1.034 0.972~1.273
小麦 Wheat	黄淮春水组 Spring-water wheat in Huanghuai region	1.013 0.997~1.036	1.127 1.068~1.182	1.123 1.066~1.179	1.121 1.065~1.181	1.034 0.963~1.242
	黄淮冬水组 Winter-water wheat in Huanghuai region	1.005 0.982~1.041	1.133 1.090~1.226	1.128 1.088~1.213	1.130 1.086~1.221	1.061 0.994~1.207
	小麦(总) Wheat (total)	1.009 0.982~1.041	1.130 1.068~1.226	1.126 1.066~1.213	1.126 1.065~1.221	1.047 0.963~1.242
	水稻(总) Rice(total)	1.007 0.974~1.043	1.095 1.009~1.294	1.096 1.008~1.298	1.084 1.009~1.287	1.068 0.974~1.414
	玉米(总) Maize(total)	1.003 0.992~1.024	1.286 1.097~1.619	1.284 1.093~1.619	1.285 1.097~1.618	1.059 0.992~1.298
	四种作物(总) Four kinds of crops(total)	1.004 0.962~1.043	1.170 1.009~1.618	1.167 1.008~1.619	1.166 1.009~1.618	1.045 0.963~1.414

从表2可以看出,BLUE估值的GF在所有区试中都接近1,总平均为1.004。这是因为BLUE即算术平均值,二者的精度是等同的。当然,BLUE的各轮GF并不正好等于1,而是在0.962~1.043间波动,这是由于交叉验证的数据分样只是所有可能分样的一部分,存在着一定的随机分样误差。一般来说,分样次数越大,误差越小。这里,分样误差造成的最大波动只有(1.043-1)=4.3%,对结果影响不大,所以本文3000次分样是足够的。如果误差过大(比如大于10%),则需增加分样次数。

表 2 中 3 种BLUP 估值的 GF 相差不大, 均平均为 1. 170 左右。这说明各种BLUP 的精度均高于算术平均值(为算术平均值的 1. 17 倍左右), 而且不同模型对其精度影响不大。相对而言, 品种和试点效应都随机的BLUPge 的精度稍高一点。不过, BLUP 的精度在不同作物间有一定差别, BLUPge 的 GF 在棉花、小麦、水稻和玉米区试中的平均分别为 1. 188、1. 130、1. 095 和 1. 286, 棉花和玉米中较高, 水稻和小麦中稍低。另外, 同一作物的不同区试组别间也有一定差别, 如黄河春棉和夏棉区试的BLUPge 的平均 GF 分别为 1. 221 和 1. 105。但是, 总的来看, 所有区试中各种BLUP 的 GF 都大于 1, 说明比起算术平均值来, BLUP 在各种区试中均有利于分析精度的提高。

最后, 表 2 中AMMI估值的平均 GF 为 1. 045, 变幅为 0. 963～ 1. 414, 而且各种区试中都比较一致; 说明虽然AMMI估值存在着比算术平均值精度高的情况, 但普遍来说, 精度提高不明显。

3　讨论

本文分析表明, BLUP 用于我国区试中品种×环境组合均值的估计, 其精度可普遍提高到目前算术平均值的 1. 17 倍左右。这意味着, 在同样精度要求下, BLUP 比算术平均值可以节约 17% 的试验小区重复数。这对我国区试精度和效率的提高来说很有意义。此外, 若结合混合线性模型求解的一些方法, BLUP 还能很好地解决不平衡数据的问题[6, 7]。目前, BLUP 的方法在作物遗传育种中正得到越来越多的关注和应用[8～13]。就本文分析结果来看, BLUP 的方法值得在我国区试中加以研究和应用。

另外, 虽然国外多数研究均表明, AMMI精度高于算术平均值(GF 最大的达 4. 30)[2, 14～18], 但本文AMMI精度的提高并不明显, 这可能与我国区试数据的自身特点有关, 譬如试点的范围和数目, 品种的数目与类型, GE 互作的强弱以及误差的大小等。这也说明, 一种统计模型或方法的精度高低, 与它所面对的数据有关。当然, 本文只是从品种×环境组合均值估计的角度探讨AMMI在我国区试中的精度特点; 至于AMMI其它方面(如品种稳定性分析)在我国区试中的应用效果及精度情况, 尚需进一步研究。

参　考　文　献

1　Peipho H P. *TAG*, 1994, 89: 647～ 654
2　Gauch H G. *Statistical Analysis of Regional Yield Trials* Elsevier, New York. 1992
3　Henderson C R. *Biometrics*, 1975, 31: 423～ 447
4　莫惠栋　农业试验统计(第二版). 上海: 上海科学技术出版社, 1992
5　Press W H, B P Flannery, et al *Numerical Recipes* Cambridge Univ Press, London. 1986
6　王松桂　线性模型的理论及应用. 合肥: 安徽教育出版社, 1987
7　朱　军　遗传模型分析方法. 北京: 中国农业出版社, 1997
8　Zhu J, B S Weir. *TAG*, 1994, 89: 160～ 166
9　Panter D M, F L Allen. *Crop Sci*, 1995, 35: 397～ 405
10　Bernardo R. *Crop Sci*, 1996, 36: 872～ 876
11　Bernardo R. *TAG*, 1997, 95: 655～ 659
12　Bernardo R. *TAG*, 1998, 97: 473～ 478
13　Bernardo R. *Crop Sci*, 1999, 39: 1277～ 1282
14　Gauch H G, R W Zobel. *TAG*, 1988, 76: 1～ 10
15　Gauch H G. *Biometrics*, 1988, 44: 701～ 705
16　Crossa J, H G Gauch, R W Zobel. *Crop Sci*, 1990, 30: 493～ 500
17　Nachit M M, G Nachit, H Ketata, et al. *TAG*, 1992, 83: 597～ 601
18　Moreno-Gonzalez J, J Crossa. *TAG*, 1998, 96: 803～ 811