

水稻品种区域试验单一性状 非平衡资料的分析方法

方秀琴 都兴林 张学君

(吉林省农科院水稻所,公主岭 136100)

提 要 目前,国内水稻品种区域试验分析只限于对产量及其性状进行方差分析和显著性测验,而对非产量性状的评价多限于根据品种的平均表现来排位,这样有可能带来误差,从而不能准确、科学地评价参试品种。而本文阐述的混合线性模型(MLM)分析原理,则可以避免该误差,从而对非平衡数据进行准确的统计分析。

关键词 水稻;区域试验;单一性状;非平衡数据;混合线性模型;分析方法

水稻品种区域试验是品种选育过程中评定参试品种和推广优良新品种的一个重要环节,通过区域试验可以获得参试品种多年、多点各农艺性状的大量数据,其中有平衡数据,也有非平衡数据。只要能对这些数据进行科学、准确的统计分析,就可以正确的评价参试品种。平衡数据(Balance Data)可用方差分析(Analysis of Variance)方法来处理。但对非平衡数据,ANOVA方法则不能对多年份的数据进行综合分析。1971年RAO提出采用分析混合线性模型(Mixed Linear Model)非平衡数据的MINQUE法(Minimum Norm Quadratic Unbiased Estimation)解决了该问题。

1 混合线性模型及其统计分析方法

1.1 混合线性模型

有 m 个品种参加多年份区域试验,各年份和各试点的区组为 r ,若每年参试的品种不完全相同,某些年份的试点也不完全相同,这就造成了多年份联合区试数据的非平衡性。则第 h 个品种在第 i 年中,第 j 个试点内第 k 个随机区组的表现型值 y_{hijk} 的混合线性模型为:

$$y_{hijk} = G_h + Y_i + L_j + YL_{ij} + GL_{ij} + GYL_{hij} + BK_{(ij)} + e_{hijk} \quad (1)$$

$$h = 1, 2, \dots, m; \quad i = 1, 2, \dots, n_h; \quad j = 1, 2, \dots, m_{hi}; \quad k = 1, 2, \dots, r;$$

其中: G_h 是品种的固定效应, $(1/m) \sum_{h=1}^m G_h = \mu$; μ 为群体的平均数; Y_i 是年份随机效应, $Y_i \sim N(0, \sigma_Y^2)$; L_j 是试点随机效应, $L_j \sim N(0, \sigma_L^2)$; YL_{ij} 是年份 \times 试点互作随机效应, $YL_{ij} \sim N(0, \sigma_{YL}^2)$; GL_{ij} 是品种 \times 试点互作随机效应, $GL_{ij} \sim N(0, \sigma_{GL}^2)$; GYL_{hij} 是品种 \times 试点互作随机效应, $GYL_{hij} \sim N(0, \sigma_{GYL}^2)$; $B_{k(ij)}$ 是年份内试点内随机区组效应, $B_{k(ij)} \sim N(0, \sigma_B^2)$; e_{hijk} 是随机误差效应, $e_{hijk} \sim N(0, \sigma_e^2)$; 第 h 个品种在 i 年中第 j 个试点内第 k 个随机区组的表型正态分布是:

$$y_{hijk} \sim N(G_h, \sigma_e^2 + \sigma_L^2 + \sigma_{YL}^2 + \sigma_{GYL}^2 + \sigma_{GL}^2 + \sigma_B^2 + \sigma_e^2) \quad (2)$$

该品种所有性状均以试点为单位统计分析。将各品种在各试点的线性模型以矩阵表示,则多年联合区试数据的混合线性模型为:

$$y = X_g + U_{yey} + U_{yLeyL} + U_{GYEGY} + U_{GLEGL} + e_{GYL} = X_g + \sum_{u=1}^5 U_u e_u + U_{reu} \quad (3)$$

其中: y 是 (nx_1) 观察值向量, y 是品种固定效应向量 ($g = [G_h]$), X 是品种固定效应的系数矩阵; e_u 是第 u 个随机变量的向量, U_u 是第 u 个随机效应的系数矩阵; $U_u = I$ 是单位矩阵。

y 的多变量多元正态分布为: $y \sim MVN(X_g \sum_{u=1}^5 \sigma_u^2 U_u V_u)$ 。

1.2 随机效应方差分量的估算

MLM 中方差分量的估算,可说明环境效应及基因型 \times 环境互作效应对性状再现的影响,该估算可直接用于估算品种效应线性对比标准误。估算方法为:用 MINQUE 法,先设方差分量检验值为常数 1,MLM 中方差分量的估计值向量 $[\sigma_u^2]$ 由 MINQUE(1) 式 ($u, \gamma = 1, 2, \dots, 6$) 解得:

$$[\text{tr}(QU_u QY)] [\sigma_u^2] = [Y^T Q U_u Q Y] \quad (4)$$

其中: tr 是矩阵的迹; 矩阵对角线元素的总和

$$U = \sum_{u=1}^5 U_u U_u^T + I$$

$$Q = U^{-1} - U^{-1} X (X^T U^{-1} X)^{-1} X^T U^{-1}$$

Y^T 是 y 的转值向量。

1.3 品种效应的线性对比及其显著性检验

MLM 中的品种固定效应 G_h 的估算方法可用最小二乘法,由品种的平均数估算。

$$G_h = y_h = (1/nh) \sum_{i=1}^{nh} [(1/nh_i) \sum_{j=1}^{nh_i} y_{hij}] \quad (5)$$

其中: nh 是品种 h 参试的年份数, nh_i 是品种 h 在第 i 年份中的试点数,参试品种线性对比可表示为:

$$C = \sum_{k=1}^k C_k y_h = C_y^T \quad (6)$$

其方差为: $\sigma^2(C) = C^T C \sum_{u=1}^5 \sigma_u^2 U_u U_u^T C$

$$\sigma^2(C) = \sum_{u=1}^5 \sigma_u^2 \quad (7)$$

用 Z 测验检验统计假设

无效假设 $H_0: \sum_{k=1}^k C_k G_h = 0$

备择假设 $H_1: \sum_{k=1}^k C_k G_h \neq 0$

如果 $C \div \sigma_{(C)}$ 的绝对值大于 $Z(\alpha)$,则可否定 H_0 ,接受 H_1 。

1.4 品种稳定性检验

对非平衡数据品种稳定性检验采用 Jackknife 法,即以每个试点的观察值作为 Jackknife 的重复抽样单位,对于回归参数的 φ_J 和抽样方差 $V(\varphi_J)$ 为:

$$\varphi_J = nk - (nk - 1)\varphi \quad (8)$$

$$V(\varphi_J) = \frac{(nh - 1)}{nk} \sum_{i=1}^{nh} \sum_{j=1}^{nh} [\varphi(ji) - \varphi]^2$$

$SE(\varphi_J) = \sqrt{V(\varphi_J)}$ 可用 t 测验对回归参数检验,用 $\varphi_J \pm t(\alpha; nh - 1) \times SE(\varphi_J)$ 确定回归参数 $100(1 - \alpha)\%$ 的置信区间。

2 实例分析

根据上述提出的统计模型分析方法,分析 1994 年吉林省水稻品种晚熟组区域试验产量资料(7 个品种,5 个点次)。由于 1994 年长春市农科院试验点缺失了一个品种吉组培 2 号

的数据,因此,本年度联合区域试验数据属于不平衡数据,可用本文提出的新方法进行统计分析。

1994年吉林省区域试验晚熟组7个品种5个点次产量分析结果如表1。

表1 省区域试验晚熟组7个品种5个点次产量分析结果 (1994年)

品 种	增产点数	减产点数	平均产量	较CK增减	位 次
			(kg/hm ²)	(%)	
秋光(CK)	0	0	8 010.9	0	6
吉组培2号	3	1	8 470.7	5.6	1
吉农大7号	4	1	8 450.7	5.4	2
吉91-2605	4	1	8 330.1	3.8	3
吉90D ₃₃	4	1	8 180.4	2.1	4
吉组培9号	4	1	8 140.4	1.6	5
九9211	1	3	7 970.9	-0.5	7

由表1可以看出,除九9211外,其它5个参试品种的产量均高于CK品种,其中以吉组培2号最高。

为了比较品种间的产量差异,可设置不同的线性对比系数(C₁, C₂, C₃, C₄, C₅, C₆, C₇)。如要比较吉90D₃₃和九9211与吉组培2号和吉91-2605的产量差异,则可设线性对比系数为(0, -1, -1, 1, 0, 0, 1),则线性对比值C = -9.16, SE = 3.10, C ÷ SE = -2.95, 其绝对值大于 $\alpha(0.05)$, 表明吉90D₃₃和九9211的产量显著低于吉组培2号和吉91-2605。

为了检验6个参试品种的稳产性,进行了稳定性分析,结果如表2。

表2 7个品种产量对环境指数回归分析的参数估计值及95%置信区间

品 种	截 距		斜 率		相关系数	
	估计值	置信区间	估计值	置信区间	估计值	置信区间
秋 光(CK)	-3.6	-12.1~5.7	1.14	0.92~1.14	0.95	0.90~0.97
吉组培2号	-1.1	-11.6~9.3	1.15	0.95~1.02	0.96	0.91~0.97
吉农大7号	-1.3	-11.0~8.4	1.21	0.97~1.20	0.97	0.92~0.98
吉91-2605	-2.1	-13.4~9.6	1.32	0.98~1.32	0.99	0.94~0.98
吉90D ₃₃	-3.5	-12.3~6.7	1.17	0.96~1.12	0.96	0.92~0.98
吉组培9号	-3.4	-12.2~6.4	1.16	0.96~1.10	0.96	0.92~0.98
九9211	-5.1	-16.1~0.9	1.26	0.97~1.20	0.98	0.96~0.99

由表2可以看出,参试品种的稳定性相似,均较稳定。

3 讨 论

本文只对产量的非平衡数据进行了分析,而对非产量性状,如品质、抗性、熟期、抽穗期等均可用上述方法进行统计分析,从而克服只根据品种的平均表现排位的方法,对品种做出科学、准确地评价。

参 考 文 献

- 1 Rao C R. Estimation of Variance and Covariance Components-MINQUE theory J. Multivar. Anal, 1971, 1
- 2 Miller R G. The Jack Knife-areview Biometrika, 1994, 6

(责任编辑:任 禾)