H. P. Piepho

Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials

Received: 5 May 1994 / Accepted: 8 September 1994

Abstract The additive main effects multiplicative interaction model is frequently used in the analysis of multilocation trials. In the analysis of such data it is of interest to decide how many of the multiplicative interaction terms are significant. Several tests for this task are available, all of which assume that errors are normally distributed with a common variance. This paper investigates the robustness of several tests (Gollob, F_{GH1}, F_{GH2}, F_R) to departures from these assumptions. It is concluded that, because of its better robustness, the F_R test is preferable. If the other tests are to be used, preliminary tests for the validity of assumptions should be performed.

Key words Genotype × environment interaction
Two-way classification · Additive main effects
multiplicative interaction (AMMI) · Cross
validation · Tests of significance · Robustness

Introduction

Genotype × environment interaction in crop cultivar trials is often analysed using the additive main effects multiplicative interaction (AMMI) model (Gauch 1988, 1992), which was originally developed in the field of social and physical sciences (Gollob 1968; Mandel 1969, 1971). The appropriate number of muliplicative interaction terms to be retained may be determined either by cross-validation (Gauch 1988; Gauch and Zobel 1988; Piepho 1994) or by significance tests. Recently, Cornelius (1993) reviewed tests of multiplicative terms for data with replication and investigated their empirical Type-I error and power via Monte Carlo simulation. He showed the F_{GH1} and F_{GH2} tests (introduced for AMMI analysis by Cornelius et al. 1992) to give satisfactory empirical Type-I errors, while the test by Gollob (1968) was too liberal when the true model contained no multiplicative terms. For the F_{GH} tests and Gollob's test it is assumed that the errors are independently normally distributed with a common variance. While in the case of proper randomization, the independence assumption is generally justified, errors may, at times, depart from the normality assumption. Moreover, it is often observed that the error variances are heterogeneous among environments. The purpose of the present paper is to investigate, via Monte Carlo simulation, the robustness of Gollob's test, the F_{GH1} and F_{GH2} tests, and the F_R test, to departures from the assumptions of normality and homogeneity of error variances.

Theory

The AMMI model for c cultivars and e environments may be written as

$$Y_{ij} = \mu + \tau_i + \delta_j + \Sigma_k \theta_k \alpha_{ik} \beta_{jk} + \varepsilon_{ij} \quad (k = 1 \text{ to } p)$$

where y_{ij} is the mean yield of the ith cultivar in the jth environment, μ is the grand mean, τ_i and δ_j are main effects of the ith cultivar and jth environment, ε_{ij} is the random error of the mean of the ith cultivar in jth environment, $p \leq \min(c-1, e-1)$. $\Sigma_k \theta_k \alpha_{ik} \beta_{jk}$ is taken to be the appropriate multiplicative model for genotype × environment interaction satisfying the constraints $\theta_1 > \theta_2 > \cdots > \theta_p > 0$, $\Sigma_i \alpha_{ik}^2 = \Sigma_j \beta_{jk}^2 = 1$, and $\Sigma_i \alpha_{ik} \alpha_{ik'} = \Sigma_j \beta_{jk} \beta_{jk'} = 0$. The multiplicative parameters are estimated by singular value decomposition (SVD) of the matrix of residuals remaining after fitting the main effects. For details regarding the estimation of model parameters see, e.g., Cornelius (1993).

The error ε_{ij} is the mean of errors ε_{ijs} of replications within an environment, i.e., $\varepsilon_{ij} = \Sigma_S \varepsilon_{ijs}/r$, where r is the number of replications per environement. Usually, it is assumed that $\varepsilon_{ij}s$ are $N(0, \sigma^2)$, where σ^2 is the variance of a cell mean. In this paper we will drop the normality assumption and investigate several nonnormal distributions for ε_{ijs}. Also, we allow for differences in environmental error variances. In this case ε_{ij} is distributed with zero mean and variance σ_j^2, where σ_j^2 is the variance of a cell mean in the jth environment.

The number of multiplicative terms appropriate for a given data set may be determined by a test of significance. The tests investigated by Cornelius (1993) are based on the statistic t_k^2/s^2, where t_k is an

Communicated by A. R. Hallauer

H. P. Piepho
Faculty of Agriculture, University of Kassel, Steinstrasse 19, 37213 Witzenhausen, Germany

estimate of the singular value θ_k, obtained by SVD, and s^2 is the pooled error mean square (f degrees of freedom $= df$) on a cell mean basis, i.e., the residual ANOVA mean square, divided by the number of replications. Three approximations to the null distribution ($H_0: \theta_k = 0$) of t_k^2/s^2 are as follows:

(1) t_k^2/s^2 is distributed as F with $(e + c - 1 - 2k)$ and f degrees of freedom (Gollob 1968).
(2) $F_{GH1} = gt_k^2/h_1 f s^2$ is distributed as F with h_1 and g df, where $h_1 = 2v_1 u_1/v_2$, $g = 2 + 2\ (f - 2)\ v_1/v_2$, $v_1 = u_2^2 + u_1^2 + (f - 4)\ u_1$, and $v_2 = (f - 2)\ u_2^2 + 2u_1^2\ u_1$ and u_2 are computed by approximations given by Cornelius (1980) for the expectation and standard deviation of the largest eigenvalue of a Wishart matrix of dimension min $(c - 1, e - 1) - k + 1$ and df max $(c - 1, e - 1) - k + 1$ (Cornelius 1993).
(3) $F_{GH2} = t_k^2/u_1 s^2$ is distributed as F with h_2 and f df, where $h_2 = 2u_1^2/u_2^2$ (Cornelius 1993).

Cornelius (1993) suggested two simulation tests based on the statistics F_{GHI} and F_{GH2}. These differ from the tests described under (2) and (3) in that u_1 and u_2 are obtained by Monte Carlo method. These tests are expected to behave very similarly to the F_{GHI} and F_{GH2}-tests based on tabular values. Because of this similarity and because of the high computational workload involved, these simulation tests are not considered here.

Another test may be performed using the residual sum of squares after fitting q multiplicative interaction terms (Cornelius, personal communication; also see Cornelius et al. 1992, who described a modified version of the F_R test applicable to the shifted multiplicative model). Under the null hypothesis that there are no more than q terms, the residual sum of squares is approximately a chi-square variable. Therefore the F-statistic

$$F_R = \left[\Sigma_i \Sigma_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 - \sum_{k=1}^{q} t_k^2 \right] \Big/ f_2 s^2$$

is approximately distributed as F with $f_2 = (e - 1 - q)\,(c - 1 - q)$ and f degrees of freedom (Gollob 1968; Goodman and Haberman 1990). When the F_R test is significant, this suggests that there is at least one more multiplicative term in addition to the q terms already fitted. Thus, the F_R test may be regarded as a test for significance of the $(q + 1)$-th multiplicative term. It has some similarity to lack-of-fit tests in linear regression. Note that for $q = 0$, i.e., when no multiplicative term is fitted, the F_R test is equivalent to the ANOVA F-test for the entire interaction, which is an exact test. Also note that the numerator df of the F_R-test is equal to the total interaction df, minus Gollob's df for the first q terms.

Simulation study

Methods

In order to investigate the robustness of the F_{GH} tests and Gollob's test to non-normality and heteroscedasticity of errors, tables of $c = 20$ genotypes and $e = 9$ environments with $r = 4$ replications were generated using the SAS procedure IML (SAS, Inc., Cary, N.C.). The dimension of the table was chosen to be comparable to simulations by Cornelius (1993), who generated tables for $c = 9$ an $e = 20$. The number of genotypes and environments was reversed, because in many cultivar evaluation trials $c \gg e$ (Piepho 1992). It is noted that the results given in Cornelius (1993) are also valid for $c = 9$ and $e = 20$.

The simulated data correspond to a completely randomized design. The RANNOR, RANUNI, RANEXP, and RANGAM functions in SAS were used to generate random deviates ε_{ijs} following, respectively, the normal, uniform, exponential, and gamma distribution. Random deviates from mixtures of two normal distributions (Cohen 1967) were generated by the RANUNI and the RANNOR functions. A description of distributions and error variances used in the simulation is given in Table 1. The distributions were scaled so that $\sigma^2 = \Sigma_j \sigma_j^2/e = \Sigma_j$ Var $(\varepsilon_{ijs})/er = 1$, where VAR$(\varepsilon_{ijs})$ denotes the variance of ε_{ijs}. With the scale-contaminated normal distributions (Distributions XII, XIII, and XVI), which mimmick the problem of outliers, this relation holds only for the mixture components, not for the mixture itself. AMMI-type interaction, subject to the usual constraints on α_{ik} and β_{jk}, was generated using the ORPOL function of SAS/IML.

Table 1 Description of distributions and error variances used in simulation

No.	Distribution	σ_j^2	$[\text{Var}(\varepsilon_{ijs}) = 4\sigma_j^2]$
(I)	Normal	1	
(II)	Normal	$0.1\,(j - 5) + 1$	
(III)	Normal	$0.2\,(j - 5) + 1$	
(IV)	Normal	0.9 for $j < e$;	1.8 for $j = e$
(V)	Normal	0.7 for $j < e$;	3.4 for $j = e$
(VI)	Uniform	1	
(VII)	Cauchy	1	
(VIII)	Exponential	1	
(IX)	Gamma (0.5)[a]	1	
(X)	$0.9\,N(0, 8/9) + 0.1\,N(0, 2)$[b]	1	
(XI)	$0.9\,N(0, 5/9) + 0.1\,N(0, 5)$[b]	1	
(XII)	$0.9\,N(0, 1) + 0.1\,N(1, 1)$[b]	1	
(XIII)	$0.9\,N(0, 1) + 0.1\,N(5, 1)$[b]	1	
(XIV)	Gamma (2)[a]	1	
(XV)	$0.95\,N(0, 10/19) + 0.05\,N(0, 10)$[b]	1	
(XVI)	$0.99\,N(0, 1) + 0.01\,N(10, 1)$[b]	1	

[a] Gamma (a) = gamma distribution with parameter a (see Johnson and Kotz 1971)
[b] $w_1 N(u, \sigma_1^2) + w_2 N(\mu, \sigma_2^2)$ = Mixture of two normal distributions with weights w_1 and w_2 (Cohen 1967)

Results

Cornelius (1993) did not investigate the F_R test. Therefore we repeated has Cases 1 to 15 for normally distributed errors. The results are shown in Table 2. The F_{GHI} test is not included because the results were identical to those of F_{GH2}. A full discussion of results for the F_{GH} tests and Gollob's tests is given in Cornelius (1993). In all cases the F_R tests had an empirical Type-I error rate close to, or below, the expected 0.05 for the θ_k values equal to zero. In this respect it was very similar to the F_{GH} tests. In most cases its power to detect the non-null θ_k terms was lower than for the F_{GH} tests. Only in Case 8 and Case 14, in which all non-null θ_k had the same value, and in Case 15 was the F_R test more powerful than the F_{GH} tests.

Results of simulation for all $\theta_k = 0$ (Case 1) and the error distributions shown in Table 1, are displayed in Table 3 (except distribution I, which is shown in Table 2). For normally distributed errors (Distribution I), Gollob's test was very liberal for the first term (Type-I error of 66%), while the F_{GH2} and F_R tests were close to the nominal error rate of 5% (see Case 1 in Table 2). The results for Gollob's test and the F_{GH2} test coincide with those by Cornelius(1993), who did not investigate the F_R test. In the other 15 cases (Distributions II to V: heteroscedasticity; Distributions VI to XVI: non-normal error distributions) the empirical Type-I error exceeded the nominal rate of Gollob's test and the F_{GH2} test, while the

Table 2 Percentage of rejections of null hypotheses in 1 000 simulated tests ($\alpha = 0.05$) of multiplicative interaction terms in 20 cultivars by nine environments tables with four replications and 13 sets of true θ_k values (Case 1 to Case 15 in Cornelius 1993). Normal distribution of errors (Distribution I)

Test	Multiplicative term (θ_k) no.																							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
	Case 1								Case 6								Case 11							
θ_k	0	0	0	0	0	0	0	0	10	10	0	0	0	0	0	0	12	10	8	0	0	0	0	0
Gollob	65.3	16.7	1.5	0.1	0.0	0.0	0.0	0.0	100.0	100.0	43.2	7.1	0.4	0.0	0.0	0.0	100.0	100.0	100.0	32.2	3.2	0.2	0.0	0.0
F_{GH2}	6.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	100.0	100.0	4.5	0.2	0.1	0.0	0.0	0.0	100.0	100.0	99.3	3.9	0.1	0.0	0.0	0.0
F_R	5.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	100.0	99.1	4.7	0.4	0.1	0.0	0.0	0.0	100.0	100.0	92.3	3.4	0.2	0.0	0.0	0.0
	Case 2								Case 7								Case 12							
θ_k	5	0	0	0	0	0	0	0	14	6	0	0	0	0	0	0	12	10	8	0	0	0	0	0
Gollob	96.4	43.8	7.1	0.4	0.1	0.0	0.0	0.0	100.0	99.0	38.2	6.6	0.4	0.0	0.0	0.0	100.0	100.0	100.0	76.7	11.0	0.3	0.0	0.0
F_{GH2}	49.5	2.5	0.1	0.1	0.0	0.0	0.0	0.0	100.0	78.2	3.2	0.1	0.0	0.0	0.0	0.0	100.0	100.0	99.7	30.9	0.9	0.0	0.0	0.0
F_R	31.7	2.2	0.2	0.1	0.0	0.0	0.0	0.0	100.0	53.8	3.5	0.4	0.0	0.0	0.0	0.0	100.0	100.0	97.8	21.0	1.2	0.1	0.0	0.0
	Case 3								Case 8								Case 13							
θ_k	10	0	0	0	0	0	0	0	5	5	5	0	0	0	0	0	50	20	10	5	0	0	0	0
Gollob	100.0	53.9	10.4	0.8	0.1	0.0	0.0	0.0	99.9	96.5	63.5	10.9	1.0	0.1	0.0	0.0	100.0	100.0	100.0	93.9	17.8	1.0	0.0	0.0
F_{GH2}	100.0	5.0	0.3	0.1	0.0	0.0	0.0	0.0	88.7	44.4	8.6	0.4	0.0	0.0	0.0	0.0	100.0	100.0	100.0	64.4	1.7	0.0	0.0	0.0
F_R	99.4	4.3	0.4	0.1	0.0	0.0	0.0	0.0	94.3	48.1	7.8	0.6	0.1	0.0	0.0	0.0	100.0	100.0	100.0	49.1	1.7	0.2	0.0	0.0
	Case 4								Case 9								Case 14							
θ_k	5	5	0	0	0	0	0	0	12	8	4	0	0	0	0	0	5	5	5	5	5	5	5	5
Gollob	99.6	82.9	23.8	3.1	0.2	0.0	0.0	0.0	100.0	100.0	82.4	20.0	1.4	0.1	0.0	0.0	100.0	100.0	99.9	98.9	88.7	55.6	17.2	1.8
F_{GH2}	75.5	20.4	1.1	0.1	0.0	0.0	0.0	0.0	100.0	99.5	25.7	0.8	0.1	0.0	0.0	0.0	100.0	98.4	89.7	66.5	37.2	17.9	4.8	1.8
F_R	72.5	15.5	1.3	0.1	0.0	0.0	0.0	0.0	100.0	98.2	18.3	1.5	0.2	0.0	0.0	0.0	100.0	100.0	99.9	98.2	83.9	50.7	14.8	1.8
	Case 5								Case 10								Case 15							
θ_k	10	5	0	0	0	0	0	0	10	10	5	0	0	0	0	0	14	6	4	4	2	0	0	0
Gollob	100.0	95.0	32.8	5.1	0.3	0.0	0.0	0.0	100.0	100.0	92.8	24.3	1.7	0.2	0.0	0.0	100.0	100.0	92.7	49.5	6.2	0.3	0.0	0.0
F_{GH2}	100.0	49.9	2.0	0.2	0.0	0.0	0.0	0.0	100.0	100.0	54.5	1.8	0.1	0.0	0.0	0.0	100.0	91.4	38.8	6.1	0.1	0.0	0.0	0.0
F_R	100.0	33.6	2.2	0.3	0.0	0.0	0.0	0.0	100.0	99.9	34.3	2.2	0.2	0.0	0.0	0.0	100.0	96.7	49.5	8.4	0.4	0.1	0.0	0.0

F_R test was very robust. The most extreme error rate for F_R (10.8% for first term) occurred for Distribution V.

Simulation results for the case $\theta_1 = 14$, $\theta_2 = 6$, $\theta_3 = \theta_4 = 4$, and $\theta_5 = 2$ (= Case 15; chosen for simlarity to the real data set analysed by Cornelius 1993) are summarized in Table 4. Under normality and homoscedasticity (Distribution I) Gollob's test for multiplicative terms two to five was superior in power, followed by the F_R test (see Case 15 in Table 2). Gollob's test was also most powerful in all other cases (Distributions II to XVI). Under heteroscedasticity (Distributions II to V), with errors from a uniform distribution (Distribution VI), and with errors from a Cauchy distribution (Distribution VII), the F_{GH} tests had better power than the F_R test, while the F_{GH} tests were more conservative with the other non-normal distributions (Distributions VIII to XVI). The power to detect the fifth multiplicative term was low for all tests and cases, while the empirical error rates for $\theta_6 = \theta_7 = \theta_8 = 0$ were within acceptable limits. The only exception was Gollob's test with Distribution VI, were the null hypothesis for the sixth term was falsely rejected in 133 of the 1 000 simulation runs.

To further investigate how power and robustness depend on the number of non-null true θ_k values, the simulation was run for errors distributed as a mixture of two normal populations, i.e., as $0.99N\ (0,1) + 0.01N(10,1)$ (Distribution XVI in Table 1). The results for cases ($\theta_1, \theta_2, \theta_3$) equal to (10, 0, 0) (Case 3), (10, 10, 0) (Case 6), and (10, 10, 5) (Case 10) are shown in Table 5. In Case 3, Gollob's test and the F_{GH} tests had a risk of 66.6% and 14.1%, respectively, of falsely declaring the first zero term (θ_2) significant, while the risk with F_R was only 3.1%. In Case 6, the Type-I error rates for the first zero term (θ_3) were 49.2% and 9% for Gollob's test and the F_{GH} tests, whereas in Case 10, the rates for the first zero term (θ_4) were 16.6% and 1.4%, respectively. So while these two tests were very liberal with regard to the first multiplicative term, they tended to be less liberal for terms two and three. In Cases 6 and 10 the F_R test was conservative, giving error rates for the first zero term of 2.7% and 0.3%, respectively. In all three cases the F_R test had less power than Gollob's test and the F_{GH} tests to detect the non-zero multiplicative terms.

Table 3 Percentage rejection of null hypotheses in 1000 simulated tests ($\alpha = 0.05$) of multiplicative interaction terms in 20 cultivars by nine environments tables with four replications. All true singular values equal to zero (Case 1). Distributions II to XVI

Test	Multiplicative term (θ_k) no.																							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
	Distribution II[a]								Distribution VII								Distribution XII							
Gollob	73.1	20.9	1.9	0.1	0.0	0.0	0.0	0.0	99.4	27.4	0.7	0.0	0.0	0.0	0.0	0.0	65.2	16.5	1.6	0.1	0.0	0.0	0.0	0.0
F_{GH2}	10.2	0.6	0.1	0.0	0.0	0.0	0.0	0.0	94.2	6.5	0.0	0.0	0.0	0.0	0.0	0.0	6.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0
F_R	5.9	0.3	0.1	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0
	Distribution III								Distribution VIII								Distribution XIII							
Gollob	84.8	33.1	3.1	0.0	0.0	0.0	0.0	0.0	73.7	20.0	2.0	0.0	0.0	0.0	0.0	0.0	68.3	17.2	1.6	0.0	0.0	0.0	0.0	0.0
F_{GH2}	25.9	1.5	0.0	0.0	0.0	0.0	0.0	0.0	10.9	0.2	0.0	0.0	0.0	0.0	0.0	0.0	7.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0
F_R	6.7	0.3	0.1	0.0	0.0	0.0	0.0	0.0	5.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	Distribution IV								Distribution IX								Distribution XIV							
Gollob	74.1	1.6	1.4	0.1	0.0	0.0	0.0	0.0	77.6	19.6	1.1	0.0	0.0	0.0	0.0	0.0	70.5	21.2	2.2	0.0	0.0	0.0	0.0	0.0
F_{GH2}	15.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	14.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	7.4	0.3	0.0	0.0	0.0	0.0	0.0	0.0
F_R	6.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0
	Distribution V								Distribution X								Distribution XV							
Gollob	93.8	3.7	0.2	0.0	0.0	0.0	0.0	0.0	66.3	17.8	1.7	0.1	0.1	0.0	0.0	0.0	79.4	20.0	1.6	0.0	0.0	0.0	0.0	0.0
F_{GH2}	63.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.2	0.2	0.1	0.0	0.0	0.0	0.0	0.0	16.1	0.4	0.0	0.0	0.0	0.0	0.0	0.0
F_R	10.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	4.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0
	Distribution VI								Distribution XI								Distribution XVI							
Gollob	63.5	16.4	2.0	0.2	0.0	0.0	0.0	0.0	73.6	18.7	1.8	0.1	0.0	0.0	0.0	0.0	86.7	26.1	1.5	0.0	0.0	0.0	0.0	0.0
F_{GH2}	4.6	0.3	0.0	0.0	0.0	0.0	0.0	0.0	10.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	22.7	0.5	0.1	0.0	0.0	0.0	0.0	0.0
F_R	4.8	0.6	0.1	0.0	0.0	0.0	0.0	0.0	4.9	0.3	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0

[a] See Table 1 for description of Distributions II to XVI. Note that Distribution I is covered by Table 2

Table 4 Percentage rejection of null hypotheses in 1000 simulated tests ($\alpha = 0.05$) of multiplicative interaction terms in 20 cultivars by nine environments tables with four replications. $\theta_1 = 14$, $\theta_2 = 6$, $\theta_3 = \theta_4 = 4$, $\theta_5 = 2$ (Cae 15). Distributions II to XVII

Test	Multiplicative term (θ_k) no.																							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
	Distribution II[a]								Distribution VII								Distribution XII							
Gollob	100.0	99.8	83.3	28.8	2.4	0.0	0.0	0.0	99.5	27.8	0.5	0.0	0.0	0.0	0.0	0.0	100.0	99.8	89.7	43.7	4.8	0.1	0.0	0.0
F_{GH2}	100.0	87.9	25.5	2.2	0.1	0.0	0.0	0.0	94.8	6.5	0.0	0.0	0.0	0.0	0.0	0.0	100.0	88.0	33.0	5.2	0.1	0.0	0.0	0.0
F_R	100.0	87.1	25.7	2.6	0.2	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	92.4	43.1	6.1	0.3	0.0	0.0	0.0
	Distribution III								Distribution VIII								Distribution XIII							
Gollob	100.0	100.0	96.8	59.2	10.3	0.3	0.0	0.0	100.0	99.9	92.5	47.5	7.0	0.2	0.0	0.0	100.0	85.8	36.7	4.9	0.3	0.0	0.0	0.0
F_{GH2}	100.0	100.0	66.0	14.5	0.5	0.0	0.0	0.0	100.0	91.8	42.3	6.3	0.5	0.0	0.0	0.0	98.5	22.7	2.0	0.0	0.0	0.0	0.0	0.0
F_R	100.0	100.0	62.8	9.8	0.5	0.0	0.0	0.0	100.0	94.9	49.0	7.6	0.3	0.0	0.0	0.0	95.9	25.1	3.3	0.2	0.0	0.0	0.0	0.0
	Distribution IV								Distribution IX								Distribution XIV							
Gollob	100.0	100.0	95.8	51.8	5.5	0.1	0.0	0.0	100.0	100.0	100.0	99.9	70.1	7.1	1.0	0.0	100.0	95.9	50.0	8.4	0.6	0.0	0.0	0.0
F_{GH2}	100.0	100.0	56.3	10.4	0.3	0.0	0.0	0.0	100.0	100.0	100.0	98.7	36.9	0.9	0.0	0.0	100.0	42.9	4.0	0.3	0.0	0.0	0.0	0.0
F_R	100.0	100.0	54.6	7.6	0.4	0.0	0.0	0.0	100.0	100.0	100.0	96.3	25.4	0.9	0.0	0.0	100.0	42.0	5.3	0.2	0.0	0.0	0.0	0.0
	Distribution V								Distribution X								Distribution XV							
Gollob	100.0	100.0	80.3	23.1	0.4	0.0	0.0	0.0	100.0	100.0	93.1	48.2	6.0	0.3	0.0	0.0	100.0	99.9	92.1	51.0	7.2	0.1	0.0	0.0
F_{GH2}	100.0	93.7	25.5	2.3	0.0	0.0	0.0	0.0	100.0	91.3	39.7	6.1	0.0	0.0	0.0	0.0	100.0	92.2	43.8	6.0	0.0	0.0	0.0	0.0
F_R	100.0	85.6	12.9	0.7	0.0	0.0	0.0	0.0	100.0	96.6	49.4	7.8	0.4	0.0	0.0	0.0	100.0	95.0	47.2	7.7	0.0	0.0	0.0	0.0
	Distribution VI								Distribution XI								Distribution XVI							
Gollob	100.0	100.0	100.0	100.0	100.0	13.3	0.6	0.0	100.0	99.9	92.9	46.9	7.7	0.0	0.0	0.0	100.9	97.0	62.6	14.5	0.7	0.0	0.0	0.0
F_{GH2}	100.0	100.0	100.0	100.0	97.9	2.3	0.1	0.0	100.0	91.5	41.1	5.9	0.2	0.0	0.0	0.0	99.9	60.0	9.7	1.0	0.0	0.0	0.0	0.0
F_R	100.0	100.0	100.0	100.0	92.0	3.1	0.3	0.0	100.0	95.8	47.8	7.6	0.4	0.0	0.0	0.0	100.0	52.4	8.2	0.8	0.0	0.0	0.0	0.0

Table 5 Percentage of rejections of null hypotheses in 1 000 simulated tests ($\alpha = 0.05$) of multiplicative interactions terms in 20 cultivars by nine environments tables with four replications and three sets of true θ_k values (Cases 3, 6, and 10 in Cornelius 1993). Errors distributed as the normal mixture $0.99\,N(0,1) + 0.01\,N(10,1)$ (Distribution XVI)

Test	Multiplicative term (θ_k) no.							
	1	2	3	4	5	6	7	8
	Case 3							
θ_k	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gollob	99.9	66.6	12.0	0.5	0.0	0.0	0.0	0.0
F_{GH2}	94.0	14.2	0.3	0.0	0.0	0.0	0.0	0.0
F_R	71.3	3.1	0.0	0.0	0.0	0.0	0.0	0.0
	Case 6							
θ_k	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
Gollob	100.0	99.5	49.2	6.7	0.1	0.0	0.0	0.0
F_{GH2}	99.1	82.7	9.0	0.4	0.0	0.0	0.0	0.0
F_R	98.1	53.0	2.7	0.1	0.0	0.0	0.0	0.0
	Case 10							
θ_k	1.0	1.0	0.5	0.0	0.0	0.0	0.0	0.0
Gollob	100.0	99.6	70.9	16.6	0.9	0.0	0.0	0.0
F_{GH2}	99.2	85.1	20.0	1.4	0.1	0.0	0.0	0.0
F_R	99.5	71.0	10.3	0.3	0.0	0.0	0.0	0.0

The robustness of cross validation (Gauch and Zobel 1988) was investigated for Cases 1 and 15 and for Distributions I, III, V, VII, XIV, XV, and XVI. Three of four replications were used for model building, while one replicate was retained for validation. Cross validation was based on the root mean squared predictive difference (RMSPD) between the model and validation data (Gauch and Zobel 1988), averaged across ten random data splittings. The model with the smallest RMSPD was taken to be the best predictive model. Simulation results are presented in Table 6. The results suggest that cross validation is robust to non-normality and to heteroscedasticity, when in fact there is no interaction (Case 1). As for Case 15, cross validation tended to detect less terms than the significance tests. Heteroscedasticity (Distributions III and V) had no serious effect on model selection, while the non-normal distributions decreased power compared to the normal case.

Table 6 Percentage of selections of multiplicative interaction terms in 20 cultivars by nine environments tables with four replications (three for model building, one for validation) in 1000 simulated cross validations (ten runs per cross validation)

Test	Multiplicative term (θ_k) no.							
	1	2	3	4	5	6	7	8
	Case 1							
.θ_k	0	0	0	0	0	0	0	0
I	4.8	1.5	0.3	0.2	0.2	0.1	0.1	0.1
III[a]	6.4	1.5	0.4	0.2	0.1	0.1	0.1	0.1
V	8.9	0.2	0.2	0.1	0.1	0.0	0.0	0.0
VII	4.3	2.5	1.6	1.1	1.0	1.0	0.9	0.7
XIV	3.8	1.0	0.4	0.3	0.2	0.1	0.0	0.0
XV	2.8	0.6	0.2	0.2	0.1	0.1	0.1	0.1
XVI	3.0	1.1	0.5	0.1	0.1	0.0	0.0	0.0
	Case 15							
θ_k	14	6	4	4	2	0	0	0
I	99.5	40.9	18.4	9.2	4.5	2.3	1.0	0.7
III	99.5	39.8	17.6	10.1	6.3	3.1	1.0	0.5
V	94.6	46.0	25.0	12.7	4.6	2.4	1.2	0.6
VII	4.4	2.5	1.7	1.3	1.0	0.9	0.6	0.5
XIV	99.0	41.6	19.8	11.0	4.6	2.2	1.2	0.3
XV	99.2	37.6	18.4	10.8	6.0	3.0	1.6	0.8
XVI	80.7	17.2	8.5	4.4	2.3	1.2	0.3	0.2

[a] See Table 1 for description of Distributions I, III, V, VII, XIV, XV, and XVI

Discussion

The F_R test is simple because it is based on a straightforward F-ratio (no tables or computation of constants needed) and the degrees of freedom are easily assigned following Gollob's rules. Furthermore, the F_R test for the first multiplicative term is very robust to non-normality and heteroscedasticity, which is not true of the F_{GH} tests. This suggests that it may be worthwhile to generally use the F_R test in place of the F_{GH} tests (and Gollob's test). If the F_{GH} tests are to be used, preliminary tests for the homogeneity of variances and for normality are in order. The simulation results indicate that the robustness of the F_R test must often be paid for by a loss in power compared to the F_{GH} tests. It is noted, however, that with an increasing number of 'true' non-zero terms, the risk of falsely declaring a term significant decreased to acceptable limits with any of the tests investigated, even if the assumptions of normality and homoscedasticity are violated.

The simulations presented in this paper were done only for 9 × 20 tables. It is conjectured (and confirmed by spot checks), however, that results for tables of other dimension are similar with regard to robustness. This conjecture needs to be checked by more extensive simulations in the future.

A simulation test similar to the one given by Cornelius (1993) could probably be devised under heteroscedasticity assumptions, though the development would not be straight forward (Cornelius, personal communication). Also, such a test would probably still be sensitive to departures from normality.

In this paper, we were mainly concerned with tests for determining how many of the multiplicative terms θ_k are non-null. As pointed out by Cornelius (1993), this is not the same issue as finding the optimal number of terms for a predictive model, which is usually done by cross validation. Often, a good predictive model has fewer terms than are judged significant by a statistical test.

Cornelius (1993) has demonstrated, however, that in some cases choosing the number of significant terms may be a better model-building strategy for prediction. Our preliminary simulations (Table 6) indicate that, although cross validation is non-parameteric in that it is not based on the normality assumption, the expected number of selected terms is not necessarily independent of the error distribution. A thorough comparison of the two model-building strategies would be worthwhile, but is beyond the scope of this paper. Because of workload limitations, we have used only ten iterations per cross validation. For an in-depth analysis, the number of iterations would probably have to be increased. Besides the number of selected multiplicative terms, a useful criterion would the interaction mean squared error (IMSE) suggested by Cornelius (1993).

Acknowledgements I thank Paul L. Cornelius (University of Kentucky) for several helpful comments, particularly for bringing the F_R test to my attention.

References

Cohen AC (1967) Estimation in mixtures of two normal distributions. Technometrics 9:15–28

Cornelius PL (1980) Functions approximating Mandel's tables for the means and standard deviations of the first three roots of a Wishart matrix. Technometrics 22:613–616

Cornelius PL (1993) Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. Crop Sci33:1186–1193

Cornelius PL, Seyedsadr M, Crossa J (1992) Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theor Appl Genet 84:161–172

Gauch HG (1988) Model selection and validation for yield trials with interaction. Biometrics 44:705–715

Gauch HG (1992) Statistical analysis of regional yield trials. AMMI analysis of factorial designs. Elsevier, New York

Gauch HG, Zobel RW (1988) Predictive and postdictive success of statistical analyses of yield trials. Theor Appl Genet 76:1–10

Gollob HF (1968) A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika 33:73–155

Goodman LA, Haberman SJ (1990) The analysis of nonadditivity in two-way analysis of variance. J Am Stat Assoc 85:139–145

Johnson NL, Kotz S (1970) Continuous univariate distributions 1. Wiley, New York

Mandel J (1969) The partitioning of interaction in analysis of variance. J Res Int Bur Stand Sect B 73:309–328

Mandel J (1971) A new analysis of variance model for nonadditive data. Technometrics 13:1–8

Piepho HP (1992) Vergleichende Untersuchungen der statistischen Eigenschaften verschiedener Stabilitätsmaße mit Anwendungen auf Hafer, Winterraps, Ackerbohnen sowie Futter- und Zuckerrüben. Doctoral Thesis (unpublished), Kiel

Piepho HP (1994) Best Linear Unbiased Prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. Theor Appl Genet 89:647–654