Rev. Brasil. Genet. 18, 1, 129-134 (1995)
(Brazil. J. Genetics)

METHODOLOGY:

Analysis of variance with interaction of effects

José B. Miranda Filho and Roland Vencovsky

ABSTRACT

The methodology given by Miranda Filho and Rissi (*Rel. Cient. Inst. Genética (ESALQ/USP) 9*: 102-114, 1975), for calculation of the sums of squares due to interactions in a simple factorial model, is presented in a more detailed fashion; an extension for a triple factorial model, with calculation of the sums of squares due to simple and triple interactions is given. Although the estimation of effects is not considered in the methodology, it may be seen as advantageous for its simplicity in performing the analysis of variance with interaction of effects. The methodology seems to be particularly useful when dealing with complex models, such as those for diallel crosses, where the analysis of variance with interaction of effects is not always possible with the usual formulas and procedures. The analysis of a variety diallel cross repeated over environments is given for illustration.

INTRODUCTION

The study of interaction of effects in the analysis of variance is an old and well known procedure. The first studies of interactions were carried out with fertilizers (factors) in combinations with varying levels of each factor. The concept was later extended to any combination of factors such as experiments for evaluation of varieties when repeated over several environments (Steel and Torrie, 1960). The analysis of variance with interaction is not too complex for simple models when the sums of squares are obtained through the usual manner. Nevertheless, for more complex models the use of matrix algebra for obtaining the sums of squares through the least squares procedure, may turn out to be a more difficult task. This is the case, for example, in the analysis of variance for diallel crosses when repeated over environments.

Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Caixa Postal 83, 13400-970 Piracicaba, SP, Brasil. Send correspondence to J.B.M.F.

Miranda Filho and Rissi (1975) showed a simple procedure for obtaining the sums of squares of the simple interaction with years of any effect in the model of a variety diallel cross; the authors also generalized the formula for different number of replications in each environment. The same procedure was used by Lima (1982) in the analysis of partial variety diallel crosses repeated over environments. Oliveira *et al*. (1987) showed all the formulas for estimation of effects and for the analysis of variance of partial diallel crosses (Miranda Filho and Geraldi, 1984) when the interactions with environmental effects are included in the model. In terms of analysis of variance, the results are essentially the same as those obtained by using the procedure of Miranda Filho and Rissi (1975). Vencovsky and Barriga (1992) also presented the formulations for the analysis of interactions, which are essentially based on the formula given by Miranda Filho and Rissi (1975). Morais *et al*. (1991) used a procedure similar to that of Oliveira *et al*. (1987) to present the formulas for estimation of effects and for the analysis of variance of complete diallel crosses. The authors showed that their results, in terms of analysis of variance, were the same when analyzed by using the expression of Miranda Filho and Rissi (1975).

The objective of this paper is to present in a more detailed fashion the methodology given by Miranda Filho and Rissi (1975) for the obtention of the sums of squares of the interaction in a simple factorial model; and to extend the methodology for simple and triple interactions in a three-way factorial arrangement. The results are restricted to calculation of the sums of squares in the analysis of variance and do not include estimation of effects. The analysis of a 5x5 variety diallel cross is given for illustration.

METHODOLOGY

Any interaction of effects in the analysis of variance comes from a two-way table in the simple factorial scheme or from n-way tables in factorials of higher orders (n).

Let us first consider the simple factorial model in a completely randomized block design:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + b_k + e_{ijk}$$

where Y_{ijk} is the value for the combination of the i^{th} level ($i = 1, 2, ..., a$) of factor A and the j^{th} level ($j = 1, 2, ..., b$) of factor B in the k^{th} replication ($k = 1, 2, ..., r$); μ is the general mean; α_i and β_j refer to the main effects of factors A and B, respectively, and $(\alpha\beta)_{ij}$ is the interaction effect; e_{ijk} is the experimental error. The analysis of variance is performed as shown in Table I.

In plant breeding, experiments repeated over environments (e.g., locations) also are analyzed in a factorial fashion. The analysis of variance is similar to that shown in Table I, with the difference that the sums of squares for replications and error are pooled over environments. The appropriate model is then:

$$Y_{ijk} = \mu + t_i + p_j + (tp)_{ij} + b_{k(j)} + e_{ik(j)},$$

where μ is the overall mean; t_i is the treatment effect; p_j is the environmental effect; $(tp)_{ij}$ is the treatment x environment interaction effect; $b_{k(j)}$ is the effect of the k^{th} replication in the j^{th} environment; and $e_{ik(j)}$ is the plot error due to the k^{th} replication of the i^{th} treatment in the j^{th} environment.

The analysis of variance with the expected mean squares is as shown in Table II.

Our purpose in this paper is to present an alternative procedure for the obtention of the interaction sum of squares. The sequence is as follows:

Table I - Analysis of variance for a two-factor arrangement in a completely randomized block design.

Source of variation	d.f.	Sums of squares
Replications	r-1	$\sum_k \frac{Y^2_{..k}}{ab} - \frac{Y^2_{...}}{rab}$
Factor A	a-1	$\sum_i \frac{Y^2_{i..}}{rb} - \frac{Y^2_{...}}{rab}$
Factor B	b-1	$\sum_j \frac{Y^2_{.j.}}{ra} - \frac{Y^2_{...}}{rab}$
Interactions AxB	(a-1)(b-1)	$\sum_{ij} \frac{Y^2_{ij.}}{r} - \sum_i \frac{Y^2_{i..}}{rb} - \sum_j \frac{Y^2_{.j.}}{ra} + \frac{Y^2_{...}}{rab}$
Error	(r-1)(ab-1)	Difference
Total	rab-1	$\sum_{ijk} Y^2_{ijk} - \frac{Y^2_{...}}{rab}$

Table II - Expected mean squares in the analysis of variance for experiments repeated over environments.

Source	d.f.	M.S.	E(M.S.)*
Replications/E	b(r-1)	M_1	--
Environments (E)	b-1	M_2	$\sigma^2 + ar\frac{\Sigma p^2}{b-1}$
Treatments (T)	a-1	M_3	$\sigma^2 + br\frac{\Sigma t^2}{a-1}$
T x E	(a-1)(b-1)	M_4	$\sigma^2 + r\frac{\Sigma (tp)^2}{(a-1)(b-1)}$
Pooled error	b(a-1)(r-1)	M_5	σ^2

*Expected mean squares for a fixed model under the completely randomized block design.

1. Calculate the sums of squares for treatment means in each environment (j); treatment means are over r replications, i.e. $\bar{Y}_{ij.} = \frac{1}{r}\sum_k Y_{ijk}$. For example, for two environments the corresponding sums of squares are:

$$S_1 = \sum_i \left(\frac{Y_{i1.}}{r}\right)^2 - \frac{(Y_{.1.}/r)^2}{a} \quad \text{and}$$

$$S_2 = \sum_i \left(\frac{Y_{i2.}}{r}\right)^2 - \frac{(Y_{.2.}/r)^2}{a}$$

2. Calculate the sum of squares for treatment means over rb observations (r replications and b environments); i.e., $S_{12} = \sum_i \left(\frac{Y_{i..}}{rb}\right)^2 - \frac{(Y_{...}/rb)^2}{a}$, where S_{12} refers to two environments; in general, for treatment means over b environments, the notation $\underline{S}$ should be used.

3. Calculate the interaction sum of squares by: $S_{TE} = rS_1 + rS_2 - 2rS_{12}$, for two environments; or

$$S_{TE} = r\sum_j S_j - rb\underline{S} \qquad \text{(I)}$$

for b environments. For different numbers of replications in each environment, the formula is:

$S_{TE} = r_1S_1 + r_2S_2 - (r_1 + r_2)S_{12}$, for two environments; or

$$S_{TE} = \sum_j r_jS_j - \left(\sum_j r_j\right)\underline{S} \qquad \text{(II)}$$

for b environments. For different number of replications in each environment, the expected mean squares are not those shown in Table II. Cochran and Cox (1957) and Bancroft (1968) provide appropriate procedures to find the coefficients of the variance components when different numbers of replications are involved.

When the treatment effect is orthogonally partitioned in the analysis of variance, the formulas for obtaining the interaction sums of squares can be applied to any of the effects in the model, and on this point lies the main advantage of the proposed methodology.

The model for three factors (say τ_i, α_j, β_k; with levels n, a and b respectively), in a completely randomized block design is:

$$Y_{ijks} = \mu + \tau_i + \alpha_j + \beta_k + (\tau\alpha)_{ij} + (\tau\beta)_{ik} + (\alpha\beta)_{jk} + $$
$$+ (\tau\alpha\beta)_{ijk} + b_s + e_{ijks}$$

where the interactions are represented in parentheses; b_s is the effect dof the s^{th} replication; and e_{ijks} is the error term. For simplicity, consider τ_i as the treatment effect which will interact with environmental effects of two sorts, α_j and β_k; then, the sums of squares for interactions involving treatment effects are calculated by:

$$S_{TA} = rb\sum_j S_j - rab\underline{S} \qquad \text{(III)}$$

$$S_{TB} = ra\sum_j S_k - rab\underline{S} \qquad \text{(IV)}$$

$$S_{TAB} = r\sum_{jk} S_{jk} - rb\sum_j \underline{S}_j - ra\sum_k \underline{S}_k + rab\underline{S} \qquad \text{(V)}$$

when considering the same number of replications in each environment; the formulas can also be adapted for different number of replications. The three-factor model also must be adapted when dealing with groups of experiments such as locations, years or seazons.

EXAMPLE

The analysis of plant height in a 5x5 variety diallel cross repeated over two locations and two plant densities is used for illustration of the proposed methodology. The experiments were in a completely randomized block design with three replications and the data for analysis are the mean of ten competitive plants per plot (Table III).

The sums of squares for each source of variation are shown in Table IV for diallel tables for each environ- ment (locations and densities) and combined over environments.

The mean squares derived from Table IV are shown in Table V.

The values of the mean squares shown in Table V are not all directly comparable because the analyses were performed with means with different numbers of replications. In fact, the individual analyses (L_1D_1 ... L_2D_2) were performed with means over r=3 replications; the combined analyses (D_1, D_2, L_1 and L_2) were performed with means over ar=br=6 replications; and the overall analyses were performed with means over abr=12 replications. The expected mean squares for the error term are then: $\sigma^2/3$; $\sigma^2/6$ and $\sigma^2/12$, respectively;

Table III - Means for plant height (m) of five varieties and ten variety crosses in two locations (L_1 and L_2) and two plant densities (D_1 and D_2).

Treatments	L_1D_1	L_2D_1	L_1D_2	L_2D_2	D_1	D_2	L_1	L_2	Overall mean*
1	1.90	1.74	1.55	1.61	1.820	1.580	1.725	1.675	1.700
2	2.17	2.25	2.15	2.33	2.210	2.240	2.160	2.290	2.225
3	2.25	2.31	2.34	2.44	2.280	2.390	2.295	2.375	2.335
4	2.15	2.31	2.31	2.41	2.230	2.360	2.230	2.360	2.295
5	2.17	2.23	2.03	2.09	2.200	2.060	2.100	2.160	2.130
1x2	2.17	2.29	2.18	2.35	2.230	2.265	2.175	2.320	2.248
1x3	2.25	2.34	2.25	2.21	2.295	2.230	2.250	2.275	2.263
1x4	2.55	2.31	1.99	2.26	2.430	2.125	2.270	2.285	2.278
1x5	2.07	2.07	1.94	2.12	2.070	2.030	2.005	2.095	2.050
2x3	2.34	2.26	2.05	2.39	2.300	2.220	2.195	2.325	2.260
2x4	2.28	2.27	2.38	2.33	2.275	2.355	2.330	2.300	2.315
2x5	2.06	2.17	2.14	2.21	2.115	2.175	2.100	2.190	2.145
3x4	2.29	2.28	2.33	2.34	2.285	2.335	2.310	2.310	2.310
3x5	2.23	2.31	2.16	2.25	2.270	2.205	2.195	2.280	2.238
4x5	1.97	2.22	2.20	2.30	2.095	2.250	2.085	2.260	2.173

D_1, D_2: averaged over locations; L_1, L_2: averaged over densities; *: averaged over locations and plant densities.

Table IV - Sums of squares in the analyses of variance of diallel tables for plant height in two locations (L_1 and L_2) and two plant densities (D_1 and D_2).

Source	L_1D_1	L_2D_1	L_1D_2	L_2D_2	D_1	D_2	L_1	L_2	Overall
Populations	0.3456	0.3156	0.6105	0.5693	0.2768	0.5506	0.3234	0.4110	0.3515
Varieties	0.0853	0.1582	0.4165	0.4001	0.1055	0.4052	0.2005	0.2624	0.2275
Heterosis	0.2603	0.1573	0.1940	0.1692	0.1713	0.1453	0.1230	0.1485	0.1239
Average het.	0.0288	0.0235	0.0247	0.0333	0.0261	0.0288	0.0267	0.0282	0.0275
Variety het.	0.1276	0.1076	0.0657	0.1196	0.1077	0.0899	0.0777	0.1065	0.0869
Specific het.	0.1039	0.0262	0.1037	0.0163	0.0375	0.0266	0.0186	0.0138	0.0096

Table V - Mean squares* in the analyses of variance of diallel tables for each environment and combined over environments.

Source	d.f.	L_1D_1	L_2D_1	L_1D_2	L_2D_2	D_1	D_2	L_1	L_2	Overall
Populations	14	2.4686	2.2540	4.3610	4.0664	1.9771	3.9328	2.3102	2.9356	2.5104
Varieties	4	2.1314	3.9554	10.4121	10.0019	2.6380	10.1312	5.0116	6.5611	5.6880
Heterosis	10	2.6034	1.5734	1.9405	1.6922	1.7127	1.4534	1.2297	1.4854	1.2394
Average het.	1	2.8830	2.3520	2.4653	3.3333	2.6107	2.8830	2.6701	2.8213	2.7452
Variety het.	4	3.1899	2.6906	1.6415	2.9905	2.6913	2.2482	1.9414	2.6623	2.1720
Specific het.	5	2.0783	0.5240	2.0747	0.3253	0.7503	0.5317	0.3722	0.2767	0.1922
Error	126**	0.7789	0.3362	0.8320	0.2895	0.2788	0.2804	0.4027	0.1564	0.1398

*Mean squares multiplied by 10^2. **From the analysis of variance with 64 treatments in three replications within each location.

where σ^2 is the error mean square in the preliminary analysis of variance (not shown in the example).

We will first consider the analysis of two locations (L_1 and L_2) under the same plant density (D_1). The interaction sum of squares for each effect in the model is calculated with data from the first, second and fifth, columns from Table IV. For example, the sum of squares for Populations x Locations is from I:

$$S_{PxL} = 3 \times 0.3456 + 3 \times 0.3156 - 6 \times 0.2768 = 0.3228$$

The same procedure is repeated for calculating the two-factor interactions for all other sources of variation.

The corresponding mean square is compatible with the preliminary analysis of variance, where the expected mean square for the error term is σ^2. In order to put together the mean squares due to the main effects and those due to interactions, the former (from Table V) must be adjusted by multiplying them by br=6 so that all the mean squares have the same expected value for the error term (σ^2). The complete analysis of variance for two locations is shown in Table VI.

For the three-factor analysis, simple and triple interactions are calculated by (for populations as source of variation):

From III: $S_{PxL} = 6 \times [0.3234 + 0.4110] - 12 \times 0.3515 = 0.1889$

From IV: $S_{PxD} = 6 \times [0.2768 + 0.5506] - 12 \times 0.3515 = 0.7467$

From V: $S_{PxLxD} = 3 \times [0.3456 + 0.3156 + 0.6105 + 0.5693] - 6 \times [0.2768 + 0.5506 + 0.3234 + 0.4110] + 12 \times 0.3515 = 0.3697$

The procedure is repeated for all other sources of variation, leading to the corresponding simple and triple interaction sums of squares. All the corresponding mean squares, calculated as above are compatible with the preliminary analysis of variance where the expected mean square for the error term is σ^2. As for Table VI, the complete analysis of variance is as shown in Table VII, where the mean squares for the main effects are obtained from Table V and adjusted by multiplying them by abr=12 in order to make them compatible with the preliminary analysis of variance.

The procedure shown in this paper for calculation of interaction sums of squares in the analysis of variance, is applicable to any model and the simplicity of the procedure is quite apparent. Oliveira *et al.* (1987) showed the procedure for the analysis of variance of partial diallel crosses according to the model of Miranda Filho and Geraldi (1984), when repeated over environments; and Morais *et al.* (1991) extended the procedure for the analysis of complete diallel crosses repeated over environments. Their methodology, in terms of analysis of variance, leads to exactly the same results as those

Table VI - Analysis of variance for plant height in two locations under the same plant density (D_1).

Source	d.f.	S.S.	M.S.	F
Populations	14	1.6608	0.11863	7.09
Varieties	4	0.6330	0.15828	9.46
Heterosis	10	1.0278	0.10276	6.14
Average het.	1	0.1566	0.15664	9.36
Variety het.	4	0.6462	0.16148	9.65
Specific het.	5	0.2250	0.04502	2.69
Pop. x L	14	0.3228	0.02305	1.38
Var. x L	4	0.0973	0.02432	1.45
Het. x L	10	0.2254	0.02254	1.35
Avg. het. x L	1	0.0004	0.00040	< 1
Var. het. x L	4	0.0597	0.01493	< 1
Spec. het. x L	5	0.1653	0.03306	1.98
Pooled error*	252	4.2151	0.01673	--

*From the preliminary analysis of variance with $E(MS) = \sigma^2$.

Table VII - Analysis of variance for plant height in two locations and two plant densities.

Source	d.f.	S.S.	M.S.	F
Locations (L)	1	0.231125	0.231125	13.78
Densities (D)	1	0.016245	0.016245	< 1
L x D	1	0.063845	0.063845	3.81
Populations	14	4.217550	0.301254	17.96
Varieties	4	2.730236	0.682559	40.69
Heterosis	10	1.487314	0.148731	8.87
Average het.	1	0.329422	0.329422	19.64
Variety het.	4	1.042554	0.260639	15.54
Specific het.	5	0.115337	0.023067	1.38
Pop. x Location (L)	14	0.188950	0.013496	< 1
Varieties x L	4	0.047207	0.011802	< 1
Heterosis x L	10	0.141743	0.014174	< 1
Avg. het. x L	1	0.000062	0.000062	< 1
Var. het. x L	4	0.062343	0.015586	< 1
Spec. het. x L	5	0.079338	0.015868	< 1
Pop. x Density (D)	14	0.746730	0.053338	3.18
Varieties x D	4	0.334359	0.083590	4.98
Heterosis x D	10	0.412371	0.041237	2.46
Avg. het. x D	1	0.000202	0.000202	< 1
Var. het. x D	4	0.014293	0.035733	2.13
Spec. het. x D	5	0.269237	0.053847	3.21
Pop. x L x D	14	0.269730	0.026409	1.57
Var. x L x D	4	0.068301	0.017075	1.02
Het. x L x D	10	0.301429	0.030143	1.80
Avg. het. x L x D	1	0.001323	0.001323	< 1
Var. het. x L x D	4	0.013669	0.003417	< 1
Spec. het. x L x D	5	0.276437	0.057287	3.42
Pooled error*	504	8.454200	0.016774	--

*From the preliminary analysis of variance with E(MS) = σ^2.

presented herein, except that the latter are not restricted to diallel crosses, as already mentioned. The estimates of the interaction effect x environment are also shown by Oliveira *et al.* (1987) and by Morais *et al.* (1991); the procedure, as proposed in this paper, is restricted to the analysis of variance and tests of hypotheses and do not foresee any estimation of effects.

ACKNOWLEDGMENTS

We are grateful to Dr. Marlene Lima (Instituto Agronômico de Campinas) for permission to use non-published experimental data in the illustrative example.

Publication supported by FAPESP.

RESUMO

A metodologia desenvolvida por Miranda Filho and Rissi (*Rel. Cient. Inst. Genética (ESALQ/USP) 9*: 102-114, 1975), para o cálculo da soma dos quadrados devida a interações em um modelo fatorial simples, é apresentada de forma mais detalhada e com extensão para um modelo fatorial triplo, com cálculos da soma dos quadrados devida a interações simples e tripla.

Embora a estimativa dos efeitos não seja considerada na metodologia, ela é vantajosa pela sua simplicidade em realizar análises de variância com efeitos de interação.

A metodologia parece ser particularmente útil quando trabalhando com modelos complexos, tais como aqueles para cruzamentos dialélicos, onde a análise de variância com efeitos de interação nem sempre é possível com as fórmulas e procedimentos usuais. A análise de um cruzamento dialélico de variedades repetidas em vários ambientes é mostrada para ilustração.

REFERENCES

Bancroft, R.A. (1968). *Topics in Intermediate Statistical Methods*. The Iowa State Univ. Press, Ames (Iowa), pp. 129.

Cochran, W.G. and **Cox, G.M.** (1957). *Experimental Designs*. 2nd edn. John Wiley & Sons Inc., New York, pp. 611.

Lima, M. (1982). Análise de cruzamentos intervarietais de milho (*Zea mays* L.) no esquema dialélico parcial. Doctoral Thesis, ESALQ/USP, Piracicaba, SP.

Miranda Filho, J.B. and **Geraldi, I.O.** (1984). An adapted model for the analysis of partial diallel crosses. *Brazil. J. Genetics 7*: 677-688.

Miranda Filho, J.B. and **Rissi, R.** (1975). Interação de efeitos genéticos com anos em um cruzamento dialélico intervarietal em milho. *Rel. Cient. Inst. Genética (ESALQ/USP) 9*: 102-114.

Morais, A.R., Oliveira, A.C., Gama, A.E.G. and **Souza Jr., C.L.** (1991). A method for combined analysis of the diallel crosses repeated in several environments. *Pesq. Agrop. Bras. 26*: 371-381.

Oliveira, A.C., Morais, A.R., Souza Jr., C.L. and **Gama, E.E.G.** (1987). Análise de cruzamentos dialélicos parciais repetidos em vários ambientes. *Brazil. J. Genetics 10*: 517-533.

Steel, R.G.D. and **Torrie, J.H.** (1960). *Principles and Procedures of Statistics*. McGraw Hill Book Co., Inc., New York, pp. 481.

Vencovsky, R. and **Barriga, P.B.** (1992). *Genética Biométrica no Fitomelhoramento*. Ed. Soc. Bras. Genética (Rib. Preto, SP), pp. 486.

(Received August 16, 1994)