Biometrical Letters
Vol. 44 (2007), No. 2, 105-128

Robustness of Joint Regression Analysis

Dulce Gamito Pereira[1], João Tiago Mexia[2], Paulo Canas Rodrigues[2]

[1] *Department of Mathematics, University of Évora, CIMA-UE (Center for Research on Mathematics and its Applications), Colégio Luís António Verney, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal*, e-mail: *dgsp@uevora.pt*
[2] *Department of Mathematics, Faculty of Sciences and Technology, Nova University of Lisbon, CMA-UNL (Center for Mathematics and its Applications), Quinta da Torre, 2825-516 Monte da Caparica, Portugal*

SUMMARY

Joint Regression Analysis is shown to be extremely robust to missing observations. Thus, using a series of "α-designs" of winter rye cultivars, it was shown that with up to 40% of missing observations the cultivars selected would be the same. In this study we considered missing observations incidences varying from 5% to 75%, with a step size of 5%. For each incidence the positions of missing observations were randomly generated in triplicate.

Key words: Joint Regression Analysis; Robustness; missing observations; Linear regressions; L_2 environmental indexes.

1. Introduction

Techniques for analysis of agricultural experiments should be robust, namely towards missing observations.

One way to study this robustness for a technique is to use a well designed set of experiments. Then for different incidences of missing observations their positions are randomly generated. Comparing the results for the complete set of experiments with those obtained from truncated sets we may find the threshold incidence rate. Above that threshold we obtain the same results than for the complete set of observations. In the case of designs for the comparison of cultivars the results are the cultivars to be selected.

Joint Regression Analysis is widely used for comparing cultivars using the results of field trials. For each block in these experiments productivity will be measured by its environmental index. Linear regressions, one per cultivar, of yields on environmental indexes are adjusted. In this process environmental indexes can be estimated by the block yield average, for complete blocks, see Mexia et al. (1997), or using a zigzag algorithm, see Mexia et al. (1999), simultaneously with the regression coefficients.

In the case of Joint Regression Analysis, Mexia et al. (1996) showed that the upper contour defined by the jointly adjusted regressions is a convex polygonal. The abscissas of the nodes in that polygonal delimit ranges in which one of the compared cultivars has the highest yields. These cultivars will be the dominant ones. For our study we will compare the lists of dominant cultivars obtained from the complete and the truncated sets of observations. Our aim is to study the robustness of Joint Regression Analysis in connection with missing observations. Thus even if a large rate of missing observations occurs in a series of experiments we can still analyze them using this technique.

We will use the data of 17 α-designs of winter rye. Each one of these designs had 4 superblocks each with 5 blocks of 4 plots. We considered incidence rates for missing observations ranging from 5% to 75% with a step size of 5%. This range covers completely the range of missing observations in the experiments to be analyzed.

2. Regression adjustment

In what follows we will assume that the yield vectors have components normally and independently distributed, so that the zigzag algorithm will lead to maximum likelihood estimators and enable us to make inferences while comparing cultivars.

Let us assume that there are b blocks and J cultivars. Representing by Y_{ij} the yield of the j^{th} cultivar, $j=1,...,J$, in the i^{th} block, $i=1,...,b$, we are led to minimize

$$S\left(\boldsymbol{\alpha}^J,\boldsymbol{\beta}^J,\mathbf{x}^b\right)=\sum_{i=1}^{b}\sum_{j=1}^{J}p_{ij}\left(Y_{ij}-\alpha_j-\beta_j x_i\right)^2,$$

where $\left(\alpha_j,\beta_j\right), j=1,...,J$, are the regression coefficients; x_i , $i=1,...,b$, are the environmental indexes and p_{ij} is the weight of the j^{th} cultivar in the i^{th} block. If the cultivar is absent we take $p_{ij}=0$. When the cultivar occurs we take $p_{ij}=p_i$, $i=1,...,b$. These weights may differ from block to block to express differences in the representativeness of the blocks. If there are several blocks in the same location, their weights will be the same. We point out that in the analysis of the data used in this paper we use *1* and *0* for weights, because we have no information indicating the relevance of the blocks which would cause us to use any other non-negative values.

The zigzag algorithm, in each iteration, first carries out a minimization in order to the regression coefficients followed by a minimization in order to the environmental indexes, hence the name of the algorithm. At the end of each iteration a standardization of the adjusted environmental indexes is carried out so that the range does not change from iteration to iteration. The procedure is carried out until the goal function stabilizes. The environmental indexes adjusted in this way are called L_2 environmental indexes, because the L_2 norm was used.

If we assume, as stated above, that the yield vectors are independent, normal, homoscedastic and that, if the j^{th} cultivar is present in the i^{th} block, the joint regression model is

$$E\left(Y_{ij}\right)=\alpha_j+\beta_j x_i\ ,\ \ i=1,\dots,b,\ j=1,\dots,J\ ,$$

and the log-likelihood will be

$$\ell\left(\boldsymbol{\alpha}^J,\boldsymbol{\beta}^J,\mathbf{x}^b\right)=-\frac{1}{2\sigma^2}S\left(\boldsymbol{\alpha}^J,\boldsymbol{\beta}^J,\mathbf{x}^b\right)-\frac{N}{2}\ln\sigma^2-\frac{N}{2}\ln\left(2\pi\right),$$

where N is the total number of yields and σ^2 is the proportionality factor for yields variance. We thus see that the zigzag algorithm leads to maximum likelihood estimators $\left(\hat{\alpha}_j,\hat{\beta}_j\right)$, $j=1,\dots,J$, of the regression coefficients and $\hat{x}_i$, $i=1,\dots,b$ of the environmental indexes.

3. α-Designs

When using incomplete blocks it is worthwhile to consider designs in which the blocks are grouped in superblocks, each containing any cultivar α times. We thus get, for all cultivars, yields obtained under similar conditions. Such designs will be resolvable in the sense of Shrikhande & Raghavarao (1963, 1964) with $\alpha=1$, and are more and more commonly used in agriculture and especially in cultivar testing. A very flexible family of resolvable designs is constituted by the α-designs introduced by Patterson & Williams (1976). While we could take $\alpha>1$, it is usually preferable to use $\alpha=1$ in order to increase the homogeneity within superblocks. There will then be a superblocks per experiment. Thus, for each cultivar, there will be a replicates per location. As we will see, this choice of $\alpha=1$ does not lead to any problems when using L_2 environmental indexes.

4. Upper contour and selection

As stated above, the upper contour defined by the adjusted linear regressions is a convex polygonal. The dominant cultivars that integrate the upper contour will have maximum yields for certain values of the environmental index. Associated to each dominant cultivar there will be a dominance range, as shown in Figure 1.

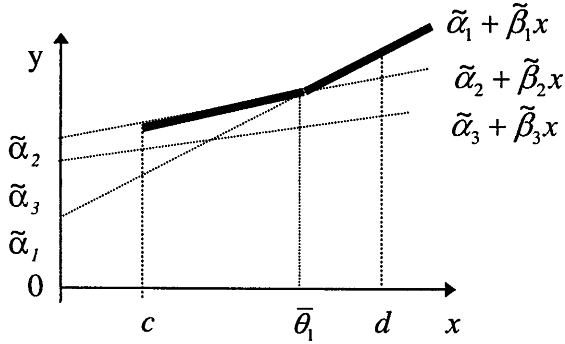

Figure 1. Upper contour with two dominant cultivars

The non-dominant cultivars should be compared with the dominant ones. To do this we begin by ordering the cultivars in decreasing order of their slopes. As we can see in Figure 2, the comparison of one non-dominated cultivar with a dominant one should be made in the left [right] extreme of the dominance range if the non-dominated cultivar has lesser [greater] slope than the dominant one.

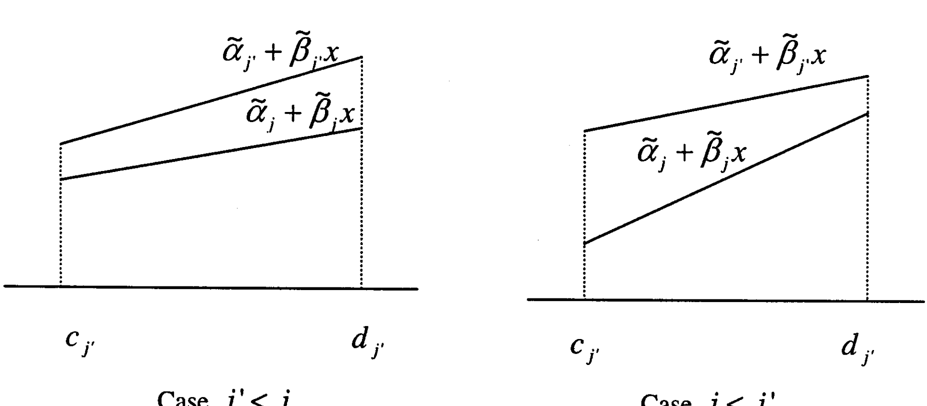

Figure 2. Comparison of cultivars

Let j' be the index of the dominant cultivar and $\lfloor c_{j'} ; d_{j'} \rfloor$ the respective dominance range. The comparison can be made using one-sided t tests. To apply this techniques we put $\mathbf{X}_j = \left[\mathbf{1} \; \vdots \; \mathbf{x}_j\right]$ with $\mathbf{x}_j$ the vector of adjusted environmental indexes for the blocks in which the j^{th} cultivar is present. We then assume that the pairs $\left(\tilde{\alpha}_j , \tilde{\beta}_j\right)$ of adjusted coefficients are normal with mean values $\left(\alpha_j , \beta_j\right)$ and covariance matrices $\sigma^2\left(\mathbf{X}_j^T\mathbf{X}_j\right)^{-1}$.

Thus, in comparing cultivars j and j', we are led to test

$$H_{\circ,j,j'}: \quad \alpha_j + \beta_j x_\circ = \alpha_{j'} + \beta_{j'} x_\circ$$

against

$$H_{1,j,j'}: \quad \alpha_j + \beta_j x_\circ < \alpha_{j'} + \beta_{j'} x_\circ ;$$

with $x_\circ$ the environmental index for which the comparison is carried out.

The t test statistics will be

$$t_{j,j'}(x_\circ) = \frac{\left(\tilde{\alpha}_{j'} + \tilde{\beta}_{j'} x_\circ\right) - \left(\tilde{\alpha}_j + \tilde{\beta}_j x_\circ\right)}{\sqrt{\frac{S}{g}\left(\begin{bmatrix}1 & x_\circ\end{bmatrix}\left(\mathbf{X}_j^T\mathbf{X}_j\right)^{-1}\begin{bmatrix}1\\ x_\circ\end{bmatrix} + \begin{bmatrix}1 & x_\circ\end{bmatrix}\left(\mathbf{X}_{j'}^T\mathbf{X}_{j'}\right)^{-1}\begin{bmatrix}1\\ x_\circ\end{bmatrix}\right)}},$$

with $S \sim \sigma^2\chi_g^2$ the sum of sums of squares of residuals, and $g = \sum^{J} b_j - 2J$, where b_j, $j = 1,\ldots,J$, is the number of blocks containing the j^{th} cultivar.

Since the $\tilde{\alpha}_j + \tilde{\beta}_j x_\circ$, $j = 1,\ldots,J$, will be normal with mean values $\alpha_j + \beta_j x_\circ$, $j = 1,\ldots,J$, and variances $\sigma^2\begin{bmatrix}1 & x_\circ\end{bmatrix}\left(\mathbf{X}_j^T\mathbf{X}_j\right)^{-1}\begin{bmatrix}1\\ x_\circ\end{bmatrix}$, $j = 1,\ldots,J$, independent between themselves and of $S \sim \sigma^2\chi_g^2$ we see that, when $H_{\circ,j,j'}(x_\circ)$ holds, $t_{j,j'}(x_\circ)$ has a central t distribution with g degrees of freedom. Thus we can use right-hand-sided t tests.

To measure the selection effectiveness, we can use the following ratios:

$$\begin{cases} r_1 = \dfrac{\textit{Number} \text{ of dominant cultivars}}{\text{Number of cultivars}} = \dfrac{L}{J} \\[2ex] r_2 = \dfrac{\textit{Number} \text{ of non dominated cultivars}}{\text{Number of cultivars}} = \dfrac{K}{J} \end{cases}.$$

5. Analysis of data without missing observations

In this application, we use the data obtained in 17 experiments in α-designs carried out by the Research Center for Cultivar Testing at Słupia Wielka (Poland) in the years 1997 and 1998.

In these experiments, cultivars of winter rye were compared. In each design, there were 4 superblocks, each one with 5 blocks of 4 plots. Each cultivar occurred on one plot per superblock.

To adjust the linear regressions we used the zigzag algorithm, and the final results are presented in Table 1.

Table 1. Adjusted coefficients and coefficient of determination.

Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	R^2
URSUS	-1.59	1.29	0.96
RAH 797	-1.60	1.22	0.97
05RAPID	-0.78	1.12	0.97
1MARDER	-0.73	1.12	0.94
RAH 897	-0.55	1.09	0.95
ESPRIT	-0.22	1.07	0.92
WID 196	-0.38	1.06	0.96
03NAD 195	-0.68	1.05	0.93
02ZDUNO	-0.82	1.02	0.97
1RAH 596	-0.15	1.01	0.95
RAH 496	0.20	1.00	0.95
1WARKO	-0.63	0.99	0.96
CHD 296	-0.55	0.98	0.93
04CHD 396	-0.54	0.98	0.95
1SMH 1195	-0.45	0.96	0.93
ADAR	-0.35	0.96	0.96
RAH 697	0.77	0.95	0.91
01AMILO	-0.27	0.93	0.93
1SMH 1295	-0.16	0.93	0.96
1SMH 1094	0.65	0.80	0.90

The adjusted regressions are presented in Figure 1.

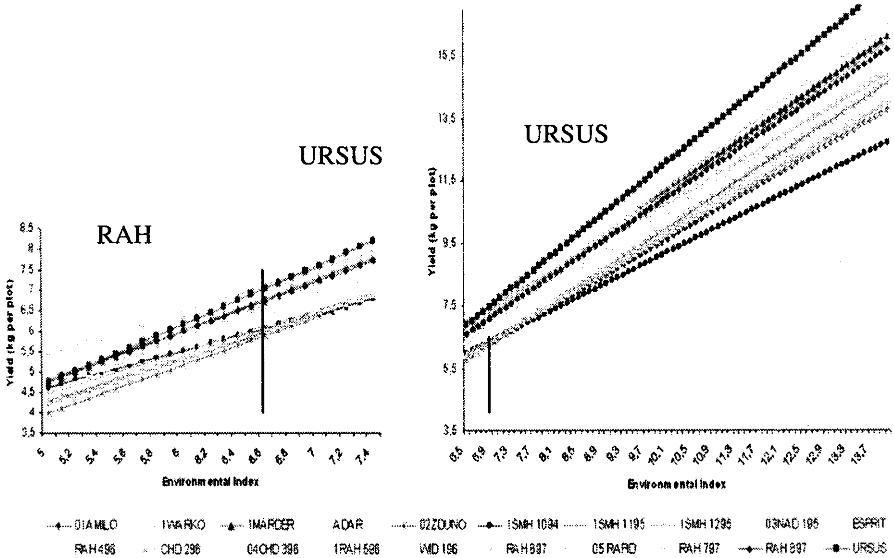

Figure 3. Adjusted regressions, using L_2 environmental indexes for the 17 experiments

Table 2 shows the results of the comparison between the cultivars which integrate the upper contour (dominants) with the remaining ones, using one-sided *t* tests.

Table 2. Dominant, dominated and non-significantly dominated cultivars

Dominant cultivars	Range of dominance	Dominated cultivars at the 5% level	Non-significantly dominated cultivars at the 5% level
RAH 697	$[5.42\ ;\ 6.84]$	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094	ESPRIT, RAH 496
URSUS	$[6.84\ ;\ 13.47]$	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094	ESPRIT, RAH 496

Table 3. Positions for discarded observations for the first replicate

Replicate 1																
36	53	105	110	120	129	136	143	146	167	195	206	215	244	273	366	372
394	428	436	446	457	458	474	489	500	594	602	608	620	632	647	668	681
694	755	760	783	790	800	825	831	843	854	875	879	899	924	950	1002	1015
1037	1040	1069	1075	1105	1115	1128	1191	1194	1217	1228	1264	1265	1318	1321	1341	1347

We can easily see that, besides the dominant cultivars (RAH 697 and URSUS), only the cultivars ESPRIT and RAH 496 are not significantly dominated at the 5% level, in the range $[5.42 \; ; \; 13.47]$. If we work at the 1% level, we will also have to consider the cultivar 1MARDER as non-significantly dominated. Thus we get the efficiency ratios $r_1 = 0.1$ and $r_2 = 0.20$ (0.25 if we work at the 1% level).

6. Analysis of data with missing observations

We decided to consider incidences of missing observations ranging from 5% to 75% with a step size of 5%. To check for robustness of the technique we randomly generated, in triplicate, the positions of the missing observations. We point out that in the generation of the positions of the missing observations we had always the care of having at least one observation per block. Since in every one of the 17 analyzed α-designs there were 4 superblocks, each with 5 blocks of 4 plots, we had 1360 observations.

6.1. Analysis with 5% incidence of missing observations. Replicate 1

The 68 randomly generated positions for discarded observations for the first replicate are given in Table 3.

As in the analysis of data without missing observations, to adjust the linear regressions we used the zigzag algorithm. The final results are presented in Table 4.

Table 4. Adjusted coefficients and coefficient of determination

Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	R^2
URSUS	-2.49	1.44	0.96
RAH 797	-2.16	1.33	0.97
1MARDER	-1.52	1.26	0.94
05 RAPID	-1.43	1.24	0.98
RAH 897	-1.24	1.21	0.95
02ZDUNO	-1.89	1.18	0.97
ESPRIT	-0.75	1.17	0.92
03NAD 195	-1.28	1.16	0.93
WID 196	-0.93	1.15	0.96
1RAH 596	-1.07	1.15	0.95

1WARKO	-1.30	1.10	0.96
04CHD 396	-1.18	1.09	0.96
1 SMH 1195	-1.19	1.08	0.94
ADAR	-1.08	1.08	0.96
CHD 296	-1.09	1.08	0.94
RAH 496	-0.09	1.07	0.95
01AMILO	-0.97	1.05	0.93
RAH 697	0.23	1.04	0.92
1 SMH 1295	-0.66	1.03	0.96
1 SMH 1094	0.14	0.90	0.90

The adjusted regressions are shown in Figure 4.

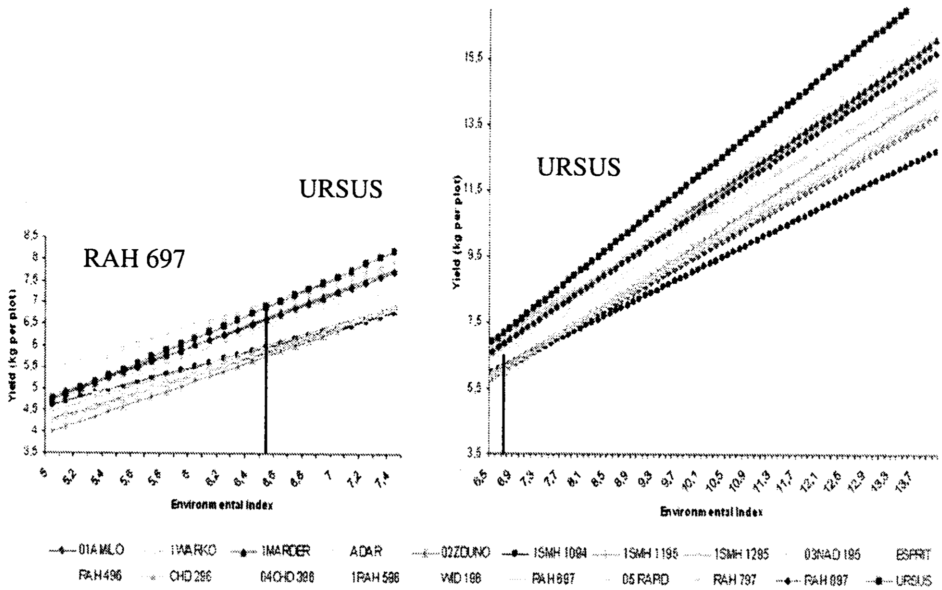

Figure 4. Adjusted regressions, using L_2 environmental indexes for the 17 experiments, when we randomly remove 5% of the observations

Table 5 shows the results of the comparison between the cultivars that integrate the upper contour (dominants) with the remaining ones, using one-sided *t* tests.

Table 5. Dominant, dominated and non significantly dominated cultivars

Dominant cultivars	Range of dominance	Dominated cultivars at the 5% level	Non-significantly dominated cultivars at the 5% level
RAH 697	$[5.43\ ;\ 6.82]$	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094	ESPRIT, RAH 496
URSUS	$[6.82\ ;\ 13.47]$	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094	ESPRIT, RAH 496

We can see that, as in the analysis of data without missing observations, besides the dominant cultivars (RAH 697 and URSUS) only the cultivars ESPRIT and RAH 496 are not significantly dominated at the 5% level, in the range $[5.43\ ;\ 13.47]$. If we work at the 1% level, we will also have to consider the cultivar 1MARDER as non-significantly dominated. Thus we get the same efficiency ratios $r_1 = 0.1$ and $r_2 = 0.20$ (0.25 if we work at the 1% level).

6.2. Analysis with 25% incidence of missing observations. Replicate 2

The 272 randomly generated positions for discarded observations for the second replicate are given in Table 6.

Table 6. Positions for discarded observations for the second replicate.

Replicate 2																
7	13	15	18	21	24	29	30	34	48	51	52	70	71	74	75	79
80	82	84	85	98	101	105	107	114	117	125	140	143	145	150	151	156
164	175	180	188	190	191	206	215	216	217	219	228	232	237	240	245	246
249	250	255	265	282	283	288	292	294	296	297	301	302	307	313	317	320
327	338	340	344	349	352	353	354	356	357	376	378	379	384	385	388	400
412	422	423	426	431	434	440	444	446	452	454	466	471	476	483	488	496
501	510	511	515	516	518	523	524	526	527	536	544	546	547	549	551	557
565	569	571	572	573	575	580	582	604	609	621	627	645	646	647	651	658
659	660	662	663	665	668	669	675	676	677	678	680	689	696	699	707	713
714	724	752	755	762	767	770	787	792	799	806	811	813	826	836	844	846
850	860	861	864	869	878	893	898	899	902	905	909	911	914	917	918	921
923	928	929	932	937	939	940	944	946	956	972	973	986	990	1000	1009	1016
1018	1019	1022	1026	1028	1033	1034	1045	1051	1053	1075	1076	1077	1079	1082	1084	1085
1093	1098	1099	1106	1107	1108	1110	1125	1127	1133	1135	1136	1139	1154	1161	1168	1171
1175	1182	1190	1191	1196	1198	1199	1207	1214	1218	1227	1240	1241	1242	1243	1246	1247
1248	1252	1269	1270	1288	1292	1294	1301	1305	1310	1316	1333	1334	1344	1356	1359	1360

As in the analysis of data without missing observations, to adjust the linear regressions we used the zigzag algorithm. The final results are presented in Table 7.

Table 7. Adjusted coefficients and coefficient of determination

Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	R^2
URSUS	-2.18	1.38	0.97
RAH 797	-1.67	1.25	0.97
05 RAPID	-0.92	1.16	0.98
WID 196	-0.97	1.15	0.96
ESPRIT	-0.58	1.14	0.93
1MARDER	-0.73	1.14	0.95
RAH 897	-0.79	1.13	0.95
RAH 496	-0.19	1.08	0.95
03NAD 195	-0.70	1.07	0.93
1WARKO	-1.25	1.07	0.97
1RAH 596	-0.36	1.06	0.96
02ZDUNO	-0.97	1.05	0.96
ADAR	-0.97	1.05	0.96
CHD 296	-0.82	1.03	0.92
1 SMH 1195	-0.80	1.02	0.93
1 SMH 1295	-0.56	1.00	0.96
01AMILO	-0.58	0.99	0.92
RAH 697	0.54	0.99	0.91
04CHD 396	-0.31	0.98	0.93
1 SMH 1094	0.59	0.84	0.90

The adjusted regressions are shown in Figure 5.

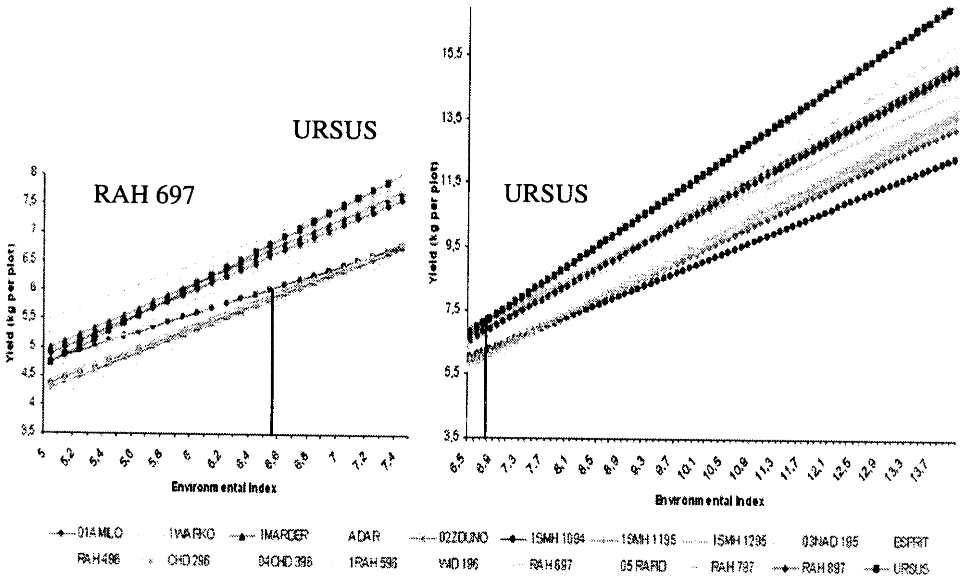

Figure 5. Adjusted regressions, using L_2 environmental indexes for the 17 experiments, when we randomly remove 25% of the observations

Table 8 shows the results of the comparison between the cultivars that integrate the upper contour (dominants) with the remaining ones, using one-sided *t* tests.

Table 8. Dominant, dominated and non-significantly dominated cultivars

Dominant cultivars	Range of dominance	Dominated cultivars at the 5% level	Non-significantly dominated cultivars at the 5% level
RAH 697	[5.43 ; 6.94]	RAH 797, 05RAPID, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094	ESPRIT, RAH 496, 1MARDER
URSUS	[6.94 ; 13.3]	RAH 797, 05RAPID, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094	ESPRIT, RAH 496, 1MARDER

We can see that besides the dominant cultivars (RAH 697 and URSUS) only the cultivars ESPRIT, RAH 496 and 1MARDER are not significantly dominated at the 5% level, in the range $[5.43\ ;\ 13.7]$. If we work at the 1% level, we will also have to consider the cultivars 05RAPID and RAH 897 as non-significantly dominated. Thus we get the efficiency ratios $r_1 = 0.1$ and $r_2 = 0.25$ (0.35 if we work at the 1% level).

6.3. Analysis with 50% incidence of missing observations. Replicate 2

As in the analysis of data without missing observations, to adjust the linear regressions we used the zigzag algorithm. The final results are presented in Table 9.

Table 9. Adjusted coefficients and coefficient of determination

Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	R^2
URSUS	-3.15	1.45	0.99
RAH 797	-2.69	1.36	0.98
1MARDER	-1.77	1.23	0.96
05 RAPID	-1.50	1.21	0.98
02ZDUNO	-2.23	1.16	0.98
RAH 897	-1.09	1.14	0.98
03NAD 195	-1.49	1.14	0.96
WID 196	-0.51	1.06	0.96
ESPRIT	-0.07	1.06	0.96
RAH 697	-0.31	1.05	0.95
1 SMH 1195	-1.17	1.03	0.95
1RAH 596	-0.38	1.03	0.98
1 SMH 1295	-1.06	1.03	0.98
ADAR	-0.78	1.02	0.97
1WARKO	-0.95	1.01	0.96
04CHD 396	-0.20	0.94	0.95
CHD 296	-0.20	0.94	0.94
1 SMH 1094	-0.54	0.92	0.93
RAH 496	1.02	0.91	0.95
01AMILO	-0.03	0.91	0.94

The adjusted regressions are shown in Figure 6.

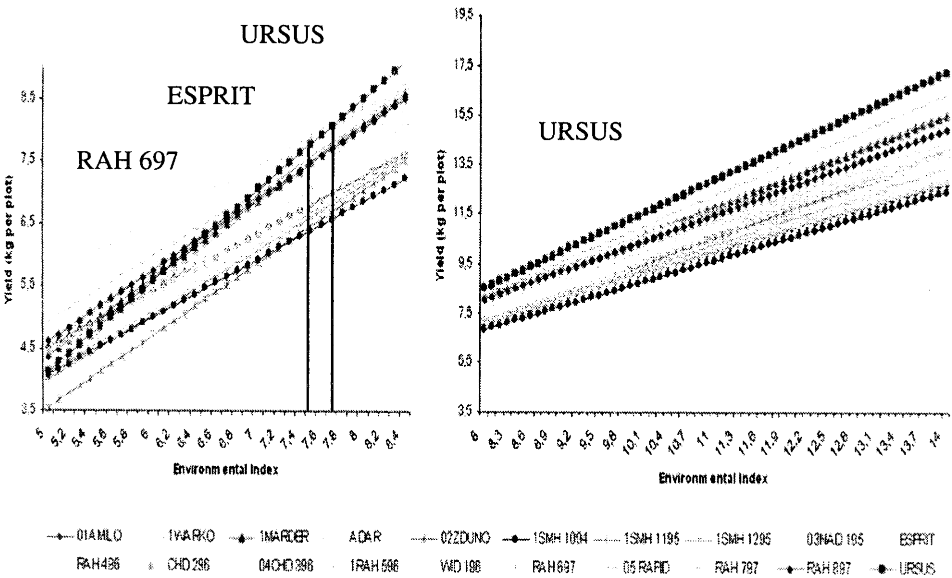

Figure 6. Adjusted regressions, using L_2 environmental indexes for the 17 experiments, when we randomly remove 50% of the observations

Table 10 shows the results of the comparison between the cultivars that integrate the upper contour (dominants) with the remaining ones, using one-sided *t* tests.

Table 10. Dominant, dominated and non significantly dominated cultivars

Dominant cultivars	Range of dominance	Dominated cultivars at the 5% level	Non-significantly dominated cultivars at the 5% level
RAH 496	[5.34 ; 7.686]	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094, RAH 697	-
ESPRIT	[7.686 ; 7.704]	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094, RAH 697	-
URSUS	[7.704 ; 13.85]	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094, RAH 697	-

We can see that we have the dominant cultivars (RAH 496, ESPRIT and URSUS) and that all the remaining cultivars are significantly dominated at the 5% level, in the range $[5.34\ ;\ 13.85]$. If we work at the 1% level, we have only to consider the cultivar 05RAPID as non-significantly dominated. Thus we get the efficiency ratios $r_1 = 0.15$ and $r_2 = 0.15$ (0.20 if we work at the 1% level).

6.4. Analysis with 75% incidence of missing observations. Replicate 1

As in the analysis of data without missing observations, to adjust the linear regressions we used the zigzag algorithm. The final results are presented in Table 11.

Table 11. Adjusted coefficients and coefficient of determination

Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	R^2
URSUS	-1.79	1.32	1.00
02ZDUNO	-2.02	1.17	1.00
RAH 496	-1.27	1.16	1.00
1MARDER	-0.61	1.13	1.00
05 RAPID	-0.49	1.10	1.00
03NAD 195	-0.89	1.09	1.00
RAH 797	-0.06	1.08	1.00
04CHD 396	-0.93	1.04	1.00
1 SMH 1195	-1.00	1.03	1.00
RAH 897	-0.02	1.02	1.00
1RAH 596	-0.02	1.01	1.00
1WARKO	-0.76	1.00	1.00
1 SMH 1295	-0.90	1.00	1.00
ADAR	-0.45	0.98	1.00
WID 196	0.53	0.96	1.00
RAH 697	0.82	0.93	1.00
CHD 296	0.01	0.92	1.00
01AMILO	1,02	0,83	1,00
ESPRIT	2,34	0,82	1,00
1 SMH 1094	0,60	0,82	1,00

The adjusted regressions are shown in Figure 7.

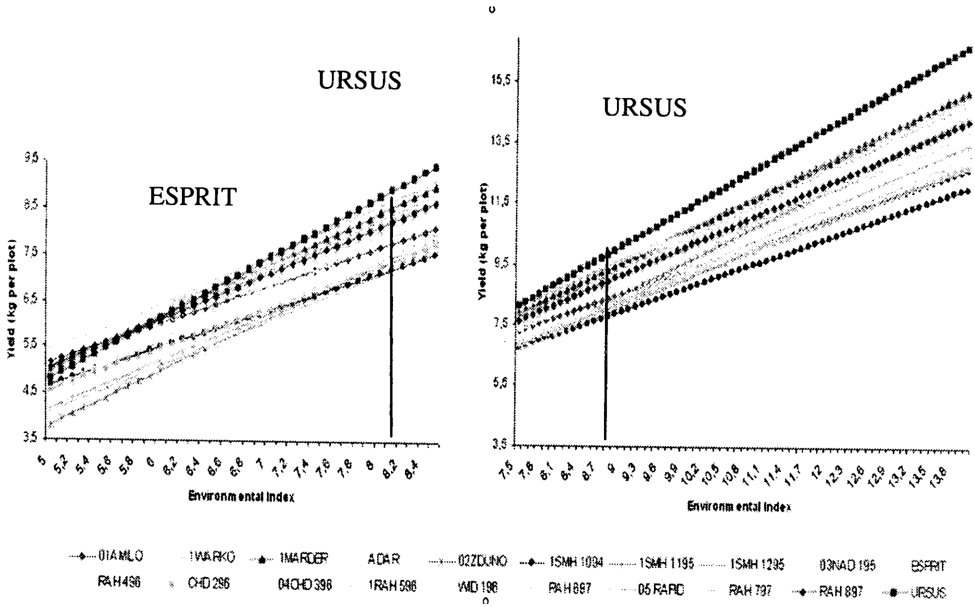

Figure 7. Adjusted regressions, using L_2 environmental indexes for the 17 experiments, when we randomly remove 75% of the observations

Table 12 shows the results of the comparison between the cultivars that integrate the upper contour (dominants) with the remaining ones, using one-sided *t* tests.

Table 12. Dominant, dominated and non-significantly dominated cultivars

Dominant cultivars	Range of dominance	Dominated cultivars at the 5% level	Non-significantly dominated cultivars at the 5% level
ESPRIT	[5.12 ; 8.24]	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094, RAH 496, RAH 697	-
URSUS	[8.24 ; 14.6]	RAH 797, 05RAPID, 1MARDER, RAH 897, WID 196, 03NAD 195, 02ZDUNO, 1RAH 596, 1WARKO, CHD 296, 04CHD 396, 1SMH 1195, ADAR, 01AMILO, 1SMH 1295, 1SMH 1094, RAH 496, RAH 697	-

We can see that we have the dominant cultivars (ESPRIT and URSUS) and that all the remaining cultivars are significantly dominated at the 5% and 1% significance levels, in the range $[5.12 \; ; \; 14.6]$. Thus we get the efficiency ratios $r_1 = 0.1$ and $r_2 = 0.1$ for both significance levels we considered.

Comment: We see that for the three replicates the dominant cultivars were URSUS and RAH 697, the same as for the experiments without missing observations. As for significantly dominated cultivars the sole difference was in replicate 2, where cultivar 1MARDER was not significantly dominated contrary to the results for the other replicates and the initial experiment.

The conclusions, at the 5% and 1% levels, for the remaining incidences of missing observations, are summarized in Table 13.

By considering these tables we discover that:

- URSUS retained its position as dominant cultivar for the most productive environments;
- Up to 40% of missing observations RAH 697 kept its position of dominance for lower environmental indexes. It was only twice superseded by RAH 496 and RAH 897. For higher rates of incidences of missing observations there is competition for that position between the cultivars RAH 697, RAH 496 and ESPRIT, with RAH 697 having the advantage;
- There were only two dominant cultivars except in single replicates for the 45%, 50% and 75% rates of incidences;
- Non-significantly dominated cultivars were almost always the same.

7. Conclusion

We can conclude that there is a good degree of stability as to dominant cultivars. This result points towards the robustness of Joint Regression Analysis in connection with missing observations and agrees with our findings – see Pereira & Mexia (2003b) – on the reproducibility of Joint Regression Analysis.

This study showed once again the good behaviour of the zigzag algorithm, once that in the case without missing observations and for all the percentages of incidences of missing observations the algorithm converges in 4 or 5 iterations.

We point out that the range of variation for adjusted environmental indexes varied a little in this study, which reinforces the robustness of the technique.

Table 13. Dominant and dominated cultivars, at 1% and 5% levels, using one-sided *t* tests

Cultivar	Significant level	Complete case	Percent of missing observations																							
			5%			10%			15%			20%			25%			30%			35%			40%		
		Replicate	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
01AMILO	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
02ZDUNO	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
03NAD	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
195	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
04 CHD	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
396	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
05RAPID	0,01	d	d	d	d	d	d	d	d	d	d			d	d	d			d		d		d	d	d		d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d		d		d	d		d
1WARKO	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1SMH	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1094	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
ESPRIT	0,01													d										d		L
	0,05													d					d		d			d		L
1RAH 596	0,01	d	d	d	d	d	d	d	d	d	d		d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
RAH 797	0,01	d	d	d	d	d	d	d	d		d	d	d	d	d	d		d	d		d	d	d	d		d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1MARDER	0,01	d				d			d	d	d	d		d	d				d	d	d	d	d	d		d
	0,05	d	d		d	d	d	d	d	d	d	d		d	d	d			d	d	d	d	d	d		d
1SMH	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1195	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
RAH 496	0,01													d			L						d		L	
	0,05										d			d	d		L					d	d	d	L	d
WID 196	0,01	d	d		d	d	d	d	d		d	d	d	d	d		d		d	d	d	d	d	d		d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d		d

RAH 897	0,01	d	d		d	d	d	d			d	d		d	d		d	d	d		d	L	d	d		d
	0,05	d	d	d	d	d	d	d	d		d	d	d	d	d	d	d	d	d		d	L	d	d		d
ADAR	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1SMH 1295	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
CHD 296	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
RAH 697	0,01	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**		**L**	**L**	**L**	**L**		**L**	**L**	**L**	**L**		**L**	**L**		
	0,05	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**		**L**	**L**	**L**	**L**	d	**L**	**L**	**L**	**L**	d	**L**	**L**		
URSUS	0,01	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**
	0,05	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**

Legend: d - Significantly dominated cultivar, at the 1% level, throughout the full range of environmental indexes;

R - Dominant cultivar at the rightmost range;

I - Dominant cultivar at an intermediate range;

L - Dominant cultivar at the leftmost range;

[shaded box] - No significantly dominated cultivar.

Table 13. (Continuation)

Cultivar	Significant level	Complete case	Percent of missing observations																				
			45%			50%			55%			60%			65%			70%			75%		
		Replicate	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
01AMILO	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
02ZDUNO	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
03NAD	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
195	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
04 CHD	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
396	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
05RAPID	0,01	d	d	d	I	d		d	L	d	d		d	d		d		d	d	d	d	I	d
	0,05	d	d	d	I	d	d	d	L	d	d	d	d	d		d		d	d	d	d	I	d
1WARKO	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	L	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	L	d
1SMH	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1094	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
ESPRIT	0,01					d	I	d		L	d		L	d		L			d	d	L	d	d
	0,05					d	I	d		L	d	d	L	d		L		d	d	d	L	d	d
1RAH 596	0,01	d	d	d		d	d	d	d	d	d		d	d		d		d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d		d	d	d	d	d	d
RAH 797	0,01	d		d	d	d	d	d	d	d	d		d	d	d	d		d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d		d	d	d	d	d	d
1MARDER	0,01		d	d	d	d	d	d		d	d	d	d	d	d	d	d	d	d	L	d	d	d
	0,05	d	d	d	d	d	d	d		d	d	d	d	d	d	d	d	d	d	L	d	d	d
1SMH	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1195	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
RAH 496	0,01			L	L	d	L	d			L		d	d	d	d	L	d	d	d	d	d	d
	0,05		d	L	L	d	L	d			L		d	d	d	d	L	d	d	d	d	d	d
WID 196	0,01	d		d	d	d	d	d	d		d		d	d		d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d		d	d		d	d	d	d	d	d	d	d

RAH 897	0,01	d	d	d		d	d	d		d	d		d	d		d		d	**L**	d	d	d	d
	0,05	d	d	d		d	d	d	d	d	d		d	d		d		d	**L**	d	d	d	d
ADAR	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d		d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
1SMH 1295	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
CHD 296	0,01	d	d	d	d	d	d	d	d	d	d	d	d	d		d	d	d	d	d	d	d	d
	0,05	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
RAH 697	0,01	**L**	**L**	**I**		**L**	d	**L**			d	**L**	d	**L**	**L**			**L**	d	d	d	d	**L**
	0,05	**L**	**L**	**I**		**L**	d	**L**			d	**L**	d	**L**	**L**			**L**	d	d	d	d	**L**
URSUS	0,01	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**
	0,05	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**

Legend: d - Significantly dominated cultivar, at the 1% level, throughout the full range of environmental indexes;

R - Dominant cultivar at the rightmost range;

I - Dominant cultivar at an intermediate range;

L - Dominant cultivar at the leftmost range;

[shaded box] - No significantly dominated cultivar.

Acknowledgements

The authors would like to thank the Research Centre for Cultivar Testing (Słupia Wielka, Poland) for supplying data used in this paper.

The first author of this work is a member of the CIMA-UE, a research center financed under the auspices of FEDER by the "Programa de financiamento Plurianual" of the Science and Technology Foundation of Portugal.

REFERENCES

Digby P. G. N. (1979). Modified joint regression analysis for incomplete variety × environment data. J. Agric. Sci., Camb., 93, 81 – 86.

Mexia J. T.; Dias G. C., Neves M. (1996). Roots of linear regressions, extensions and applications to joint regression analysis. Trabalhos de Investigação, nº1. FCT/UNL.

Mexia J. T., Amaro A. P., Gusmão L., Baeta J. (1997). Upper contour of a Joint Regression Analysis. J. Genet. & Breed. 51, 253-255.

Mexia J. T., Pereira D. G., Baeta J. (1999). L_2 environmental indexes. Biometrical Letters 36, 137-143.

Mexia J. T., Pereira D. G. (2001). Joint regression analysis for winter rye cultivars using L_2 indexes. Colloquium Biometryczne 31, 207-212.

Ng M. P., Grunwald G. K. (1997). Nonlinear regression analysis of the joint-regression model. Biometrics, 53, 1366 – 1372.

Ng M. P., Williams E. R. (2001). Joint-regression analysis for incomplete two-way tables. Aust. N. Z. J. Stat., 43(2), 201 – 206.

Patterson H. D., Williams E. R. (1976). A new class of resolvable incomplete block designs. Biometrika, 63, 83 – 92.

Pereira D. G., Mexia J. T. (2002). Multiple comparison in Joint Regression Analysis with special reference to variety selection. Scientific papers of the Agricultural University of Poznan, Agriculture 3, 67-74.

Pereira D. G., Mexia J. T. (2003a). The use of Joint Regression Analysis in selecting recommended cultivars. Biuletyn Oceny Odmian (Cultivar Testing Bulletin) 31, 19-25.

Pereira D. G., Mexia J. T. (2003b). Reproducibility of Joint Regression Analysis. Colloquium Biometryczne 33, 279-291.

Scheffé H. (1959). The Analysis of Variance. John Willey & Sons, New York.

Seber G. A. F. (1977). Linear Regression Analysis. John Willey & Sons, New York.

Shrikhande S. S, Raghavarao D. (1963). A method of construction of incomplete block designs. Sankhya Ser. A 25, 399 – 402.

Shrikhande S. S, Raghavarao D. (1964). Affine α-resolvable incomplete block designs. In, C. R. Rao (ed.), Contributions to Statistics. Pergamon Press, Statistical Publishing Society, Calcutta, 471 – 480.