Listy Biometryczne - Biometrical Letters
Vol. 36(1999), No. 2, 137-143

L_2 environmental indexes*

João Tiago Mexia[1], Dulce Pereira[2] and José Baeta[3]

[1]Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2825 Monte da Caparica, Portugal;
[2]Universidade de Évora, 7000 Évora, Portugal;
[3]Estação Agronómica Nacional, 2780 Oeiras, Portugal

SUMMARY

The use of L_2 norm in defining environmental indexes is discussed. For each type of L_2 indexes a different type of regression of yield on environmental index is considered. The indexes are chosen in order to minimize the sum of the sums of squares of residuals for the regressions, there being a regression per variety. The special cases of linear and cubic indexes are singled out as well as the general polynomial case. The first of these cases constitutes an alternative to the classical Joint Regression Analysis in which the environmental indexes of the different blocks are their average yields.

KEY WORDS: L_2 norm, environmental indexes, joint regression analysis.

1. Introduction

Joint Regression Analysis has been widely used, see Mooers (1921), Finlay and Wilkinson (1963) and Eberhart and Russel (1966), in comparing cultivars. Usually there is a field network of random block designs, the environmental index for each block being estimated, see Gusmão (1985), by it's average yield. Linear regressions on these indexes are then adjusted for the varieties. Recently, Mexia et al. (1997) proposed that the environmental indexes for blocks should minimize the sum of the sums of squares of the residuals. These indexes are already L_2 indexes since L_2 norm is used. Despite they having been named quadratic indexes we now will refer to them as linear L_2 indexes. In this way we are placing then inside a general class of environmental indexes defined through the use of L_2 norm. Each type of these indexes will correspond to a type of regression. In what follows we will consider, first the linear, then the cubic and lastly the general polynomial indexes.

*The paper was submitted on the occasion of 70-th birthday of Professor Tadeusz Caliński.

2. Linear L_2 indexes

Let the j-th cultivar have yield y_{ij} in the i-th block, $i = 1, ..., n$, $j = 1, ..., J$. We intend to obtain environmental indexes θ_i, $i = 1, ..., n$, such that the sum of the sums of squares of residuals for the corresponding linear equations is minimal.

If we had the indexes vector $\mathbf{x} = (x_1, ..., x_n)^T$ we would have the sum of sums of squares of residuals

$$S(\mathbf{x}) = \sum_{j=1}^{J}\sum_{i=1}^{n}(y_{ij} - y_{0j})^2 - \sum_{j=1}^{J}\frac{\left(\sum_{i=1}^{n}(y_{ij} - y_{0j})(x_i - x_0)\right)^2}{\sum_{i=1}^{n}(x_i - x_0)^2}, \tag{1}$$

where $y_{0j} = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$ and $x_0 = \frac{1}{n}\sum_{i=1}^{n} x_i$. Thus we will have to maximize

$$h(\mathbf{x}) = \sum_{j=1}^{J}\frac{\left(\sum_{i=1}^{n}(y_{ij} - y_{0j})(x_i - x_0)\right)^2}{\sum_{i=1}^{n}(x_i - x_0)^2}. \tag{2}$$

Since $h(a\mathbf{x}+b\mathbf{1}) = h(\mathbf{x})$ we can restrict ourselves to vectors $\mathbf{x}$ such that $\sum_{i=1}^{n} x_i^2 = 1$ and $\sum_{i=1}^{n} x_i = 0$. It then may be shown, see Mexia et al. (1997), that the solution is the first eigenvector $\boldsymbol{\alpha}_1$ of the matrix $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T$ with $\tilde{\mathbf{Y}} = [\tilde{y}_{ij}]$ and $\tilde{y}_{ij} = y_{ij} - y_{0j}$; $i = 1, ..., n$; $j = 1, ..., J$. In most applications n exceeds J and it is convenient to start by obtaining the first eigenvector γ_1 of $\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}$ since it is straightforward to show that the corresponding eigenvalue λ_1 is the same for both eigenvectors, one of $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T$ and the other of $\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}$, and that

$$\boldsymbol{\alpha}_1 = \frac{1}{\sqrt{\lambda_1}}\tilde{\mathbf{Y}}\gamma_1. \tag{3}$$

Once the linear L_2 indexes are adjusted we can rescale them so that the minimum and maximum linear L_2 indexes are equal to the minimum and maximum classical indexes.

In the field network we are going to consider, there were eleven four random block designs and the cultivars: CELTA, HELVIO, TE9006, TE9007, TE9008, TE9110, TE9115, TE9204 and TROVADOR were compared. In Table 1 we present the results obtained using the classical and the rescaled linear L_2 indexes.

The sums of squares of residuals were 2147355764 for the classical and 21292241572 for the linear L_2 indexes.

In Figures 1a and 1b we present the two sets of adjusted regressions.

In each case the upper contour of the regression lines is a polygonal, the cultivars whose regressions belong to the upper contour being dominant. Ordering the cultivars according to their decreasing slopes let the i-th cultivar, with adjusted regression $\alpha_i^* + \beta_i^* x$, be dominant in range $[c_i; d_i]$. Now, see Mexia et al. (1997), we can carry

Table 1. Adjusted regressions

Cultivar	Classical indexes			Linear L_2 indexes		
	a^*	β^*	R^2	a^*	β^*	R^2
CELTA	–462.19476	1.22745	0.90745	–538.25422	1.24364	0.91838
HELVIO	–101.95424	1.04521	0.92415	–148.15485	1.05487	0.92801
TE9006	–351.26111	1.09817	0.86267	–409.17417	1.11040	0.86953
TE9007	–480.25999	1.10783	0.91716	–537.35932	1.11987	0.92397
TE9008	411.56732	0.94378	0.89990	380.94742	0.95004	0.89898
TE9110	–42.59616	0.88260	0.78279	–32.64602	0.87988	0.76698
TE9115	1166.65400	0.56949	0.54980	1227.34924	0.55568	0.51606
TE9204	67.32710	1.06924	0.89540	20.56182	1.07901	0.89895
TROVADOR	–207.28216	1.05622	0.89297	–249.08800	1.06490	0.89487

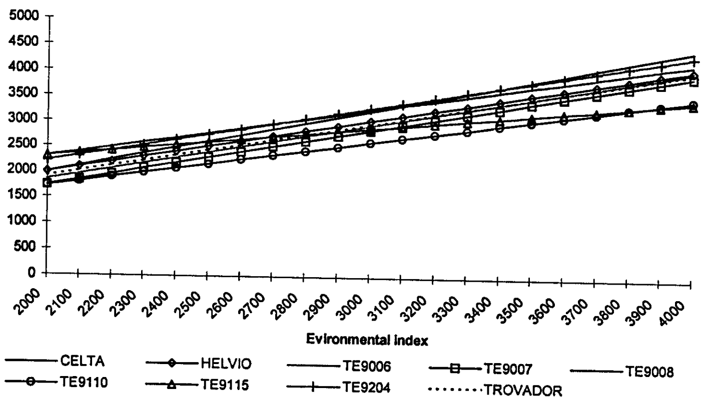

Figure 1a. Joint linear regression with classical indexes

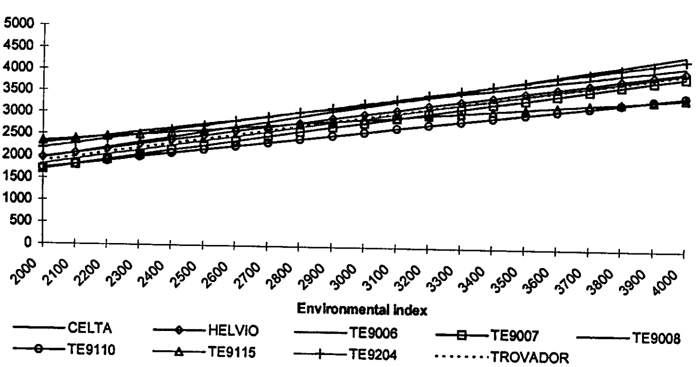

Figure 1b. Joint linear regression with L_2 indexes

out right-sided t-tests to compare this dominant cultivar with the others. Let $x_1, ..., x_n$ be the environmental indexes (classical or L_2 linear) and $S_1, ..., S_J$ the sums of squares of residuals, with $x_0 = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s_{xx} = \sum_{i=1}^{n}(x_i - x_0)^2$ and

$$k(u) = \frac{1}{n} + \frac{(x_0 - u)^2}{s_{xx}}. \tag{4}$$

The test statistics are

$$\begin{cases} t_{il} = \dfrac{(\alpha_i^* + \beta_i^* c_i) - (\alpha_l^* + \beta_l^* c_i)}{\sqrt{k(c_i)\dfrac{(S_i + S_l)}{n-2}}}; & \beta_l^* < \beta_i^*, \\ t_{il} = \dfrac{(\alpha_i^* + \beta_i^* c_i) - (\alpha_l^* + \beta_l^* c_i)}{\sqrt{k(c_i)\dfrac{(S_i + S_l)}{n-2}}}; & \beta_l^* < \beta_i^*, \end{cases} \tag{5}$$

and are based on $2(n-2)$ degrees of freedom. In the case of our exemplary data this technique led us to results given in Table 2.

Thus the L_2 linear indexes present a slightly better separation power.

Table 2. Dominant and dominated cultivars

Classical indexes			Linear L_2 indexes		
Dominant	Dominance range	Dominated	Dominant	Dominance range	Dominated
CELTA	[3346.8399; 8838.4922]	TE9007 TROVADOR TE9110	CELTA	[3394.3623; 8838.4922]	TE9007 TE9006 TROVADOR TE9110 TE9115
TE9204	[2743.8491; 3346.8399]	TE9007 TE9006 TE9110	TE9204	[2794.2819; 3394.3623]	TE9007 TE9006 TE9110
TE9008	[2212.5922; 2743.8491]	TE9007 TE9110	TE9008	[2212.5922; 2794.2819]	TE9007 TE9110

3. Polynomial L_2 indexes

3.1. Adjustment

We now adjust polynomial regressions. If $k-1$ is their degree we will have k coefficients per regression.

With $\mathbf{x} = (x_1, ..., x_n)^T$ let $\mathbf{x}^{(j)} = (x_1^j, ..., x_n^j)^T, j = 0, 1, ...$, and $\Omega_k(\mathbf{x})$ be the range space of $\left[\mathbf{x}^{(0)} \vdots \dots \vdots \mathbf{x}^{(k-1)}\right]$. Given the vectors $\mathbf{y}_1, ..., \mathbf{y}_J$ of cultivar yields we want to minimize the sum of the sums of squares of the residuals for polynomial regressions. We point out that we are minimizing a function of the vector $\mathbf{x}$ of environmental indexes, that for, each such vector $\tilde{\mathbf{x}} = (\tilde{x}_1, ..., \tilde{x}_n)^T$, the adjusted coefficient vectors are

$$\tilde{\boldsymbol{\beta}}_j(\tilde{\mathbf{x}}) = \left(\mathbf{X}_k(\tilde{\mathbf{x}})^T \mathbf{X}_k(\tilde{\mathbf{x}})\right)^{-1} \mathbf{X}_k(\tilde{\mathbf{x}})^T \mathbf{y}_j,\ j = 1, ..., J, \tag{6}$$

and that the sum of the sums of squares of residuals is

$$\begin{aligned} S_k(\tilde{\mathbf{x}}) &= \sum_{j=1}^{J} \sum_{i=1}^{n} \left(y_{ij} - \sum_{l=1}^{k} \tilde{\beta}_{jl}(\tilde{\mathbf{x}}) \tilde{x}_i^{l-1} \right)^2 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{J} \left(y_{ij} - \sum_{l=1}^{k} \tilde{\beta}_{jl}(\tilde{\mathbf{x}}) \tilde{x}_i^{l-1} \right)^2 . \end{aligned} \tag{7}$$

This last expression suggests an algorithm to carry out the minimization of $S_k(\tilde{\mathbf{x}})$. We start by taking as environmental indexes vector the vector of classical indexes $\tilde{\mathbf{x}}_1$ and obtain the $\tilde{\beta}_j(\tilde{x}_1)$, $j = 1, ..., J$. Next we minimize separately

$$\sum_{j=1}^{J} \left(y_{ij} - \sum_{l=1}^{k} \tilde{\beta}_{jl}(\tilde{x}_1) \tilde{x}_{1i}^{l-1} \right)^2$$

treating it as a function of $x_{1,i}$, $i = 1, ..., n$. We thus obtain the components of a new vector which we rescale in order to have the same range for the environmental indexes. With $\mathbf{x}_2$, the rescaled vector, we enter the second iteration and repeat this procedure till the sums of squares of residuals stabilize.

In order to avoid numerical problems it may be convenient to substitute the $\mathbf{y}_j$, $j = 1, ..., J$ by the $\tilde{\mathbf{y}}_j = \frac{1}{q}\mathbf{y}_j$, $j = 1, ..., J$, so that all components are in the range $[0; 1]$. Once the adjustment is carried through the final sum of squares of residuals has to be multiplied by q^2 and the l-th components of the adjusted coefficient vectors by q^{2-l}, $l = 1, ..., k$.

Since $\Omega_k(a\mathbf{x}+b\mathbf{1}) = \Omega_k(x)$ we could restrict ourselves to vectors with $\sum_{i=1}^n x_i^2 = 1$ and so, according to the Weierstrass theorem, $S_k(\tilde{\mathbf{x}})$ has always a minimum.

3.2. Model validation

The possibility of adjusting environmental indexes of increasing degrees enables us to validate the choice of degree. Namely, it may be interesting to validate the use of linear quadratic indexes since these are easy to interpret. To validate the use of $h-1$ degree indexes we can, with $h < k$, carry out adjustments both for the $h-1$ and the $k-1$ degree indexes. For the validation of the linear indexes, $h = 2$, we may consider taking $k = 4$ since, as an alternative to a linear response to the environmental index, we could think of a response curve with a range of increasing returns followed by a decreasing returns range. Both ranges would be separated by an inflexion point and $k - 1 = 3$ is the lowest degree for which such points exist.

Let $\tilde{\mathrm{x}}$ be the vector of the k degree adjusted environmental indexes. To test if the last $k - h$ coefficients are null for all cultivars we have the F test with statistic

$$F = \frac{J(n-h)}{J(k-h)} \frac{S_h(\tilde{\mathrm{x}}) - S_k(\tilde{\mathrm{x}})}{S_k(\tilde{\mathrm{x}})} \tag{8}$$

with $J(k-h)$ and $J(n-h)$ degrees of freedom.

In the example we are considering we get $S_2(\tilde{\mathrm{x}}) = 21292241572$, $S_4(\tilde{\mathrm{x}}) = 17440482366$ and $F = 4.42$ with 8 and 360 degrees of freedom. Thus F is very highly significant and it may be worthwhile to pursue the study as we are presently doing.

References

Eberhart, S.A., Russel, W.A. (1966). Stability parameters for comparing varieties. *Crop Sci.* **6**, 36-40.

Finlay, K.W., Wilkinson, G.N. (1963). The analysis of adaptation in a plant breeding programme. *Aust. J. Res.* **14**, 742-754.

Gusmão, L. (1985). An adequate design for regression analysis of yield trials. *Theor. Appl. Genet.* **71**, 314-319.

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., Baeta, J., Pereira, D. (1997). Quadratic environmental indexes. International Symposium on Rice Quality, 24-27 November, Nottingham, UK.

Mooers, C.A. (1921). The agronomic placement of varieties. *J. Amer. Soc. Agron.* **13**, 337-352.

Received 5 September 1998; revised 10 April 1999

Wskaźniki środowiskowe typu L_2

STRESZCZENIE

Dyskutowane jest wykorzystanie normy L_2 do definiowania wskaźników środowiskowych. Dla każdego typu wskaźnika rozważa się inny typ regresji względem środowiska. Wskaźniki są dobrane tak aby minimalizowały sumę sum kwadratów dla błędów w regresji dla poszczególnych odmian. Wyróżniono szczególne przypadki wskaźników liniowych i kwadratowych oraz ogólny typ wskaźników wielomianowych. Pierwszy z tych przypadków stanowi alternatywę dla klasycznej regresji w której wskaźnikami środowiskowymi są średnie plony odmian w środowiskach.

SŁOWA KLUCZOWE: norma L_2; wskaźniki środowiskowe, analiza wspólnej regresji