

IOLA MARIA SILVÉRIO PINTO

ANÁLISE CONJUNTA DE REGRESSÕES
E PLANOS DE MELHORAMENTO

Dissertação apresentada para obtenção do
Grau de Doutor em Matemática-Estatística
pela Universidade Nova de Lisboa, Facul-
dade de Ciências e Tecnologia.

Lisboa

2006

Agradecimentos

Ao longo do tempo de realização deste trabalho, aprofundi e consolidei conhecimentos adquiridos, realizei pesquisa, estudei e assimilei novos conceitos, desenvolvi os meus raciocínios tentando resolver um problema determinado e assim foi surgindo esta tese. Ao mesmo tempo em que fiz tudo isto, recebi do Professor João Tiago Mexia, meu supervisor, uma grandiosa orientação, manifestada permanentemente por novas explicações, sugestões e críticas essenciais à realização desta tese. Para ele, o meu infinito agradecimento, não só pelo seu permanente e atencioso acompanhamento, mas também pela dedicação e encorajamento demonstrados nos momentos em que disso precisei.

A família é um pilar essencial na nossa vida. Acompanharam-se sempre durante este longo percurso, sabendo escutar-me e aconselhar-me nos momentos mais difíceis. Assim é com Amor, carinho e estima que agradeço e reconheço todo o apoio, compreensão e paciência que recebi dos meus pais, do meu marido Armando e do meu filho Diego, que começou a ouvir falar destas coisas ainda era um pequenino embrião e tem tido, desde então, uma mãe com uma disponibilidade muito limitada. *É com muita gratidão que realço o profundo agradecimento à minha mãe, por ter sido mãe "duas vezes", ao longo de todo este período.* Gostaria ainda de mencionar todo o encorajamento e motivação que sempre recebi dos meus avós e o apoio da minha irmã Inês.

Agradeço também a todos os meus amigos que me deram força e sempre me encorajaram para seguir em frente. Em especial, à minha amiga Célia Nunes, que me ajudou sempre oferecendo a sua grande amizade, sobretudo ouvindo-me sempre que precisei. Não posso esquecer também todas as informações que ela, a Sandra e o Dario me deram sobre o funcionamento do Latex.

Quero ainda dedicar os meus agradecimentos sinceros às seguintes pessoas:

- Ao professor Stanislaw Mejza pelas críticas construtivas, sugestões e explicações;
- Ao Professor Tadeusz Calinski pelas suas importantes sugestões e comentários;
- Às pessoas do Departamento de Cereais da Estação Nacional de Melhoramento de Plantas, nomeadamente ao Senhor Engenheiro Bemvindo Maçãs e ao Senhor Engenheiro João Coco, pela recolha, envio e explicação dos dados utilizados na aplicação prática desta tese;

Por último deixo também o meu agradecimento às entidades que contribuíram para a atribuição da bolsa de estudo que me permitiu dedicar-me inteiramente a este trabalho.

Resumo

A Análise Conjunta de Regressões, ACR, é uma técnica conhecida pela sua utilização na comparação e selecção de cultivares. A sua aplicação estava geralmente condicionada a ensaios realizados num número reduzido de anos, para um conjunto fixo de cultivares. Neste trabalho, vamos estudar a extensão do âmbito de aplicação desta técnica a planos de melhoramento.

Os ensaios analisados através da ACR encontram-se divididos em blocos. A produtividade de cada bloco é medida pelo respectivo índice ambiental. A técnica da ACR, como o seu próprio nome indica, consiste no ajustamento de regressões lineares do rendimento de cada cultivar no respectivo índice ambiental.

Numa primeira parte, apresentamos as técnicas estatísticas que serão utilizadas posteriormente na análise dum plano de melhoramento concreto. Mostraremos nomeadamente, como realizar os ajustamentos das regressões lineares e como proceder à comparação e selecção dos cultivares.

Na segunda parte o nosso trabalho mostrará, revendo-se no Plano de Melhoramento de Trigo Mole em Portugal (1986 - 2000), como utilizar a ACR em planos de melhoramento.

Apresentaremos ainda um modelo para os índices ambientais, usando variáveis dummy associadas a anos, locais e cultivares. Os coeficientes ajustados para os anos foram usados para expressar, no caso do Programa de Melhoramento do Trigo Mole em Portugal, os efeitos atribuíveis ao factor ano.

Os resultados obtidos no caso Português, indicam claramente a utilidade da técnica da ACR no acompanhamento de programas de melhoramento, mais precisamente, na tomada de decisão relativa aos cultivares que devem permanecer no plano de melhoramento de um ano para o outro.

Abstract

Joint Regression Analysis is a well known technique for cultivar comparison and selection. Usually, a small number of years and a given list of cultivars are considered, when applying the technique. We now discuss the extension of the technique to plant breeding programs.

The field trials analyzed through JRA are divided in blocks. The productivity of each block is measured by its environmental index. Per cultivar, a linear regression of the yield on environmental index is adjusted, which gave the technique its name.

In a first part, the required statistical techniques are presented. Namely, it is shown, how to adjust the linear regressions and how to compare cultivars.

Nextly, the application of JRA to plant breeding programs is performed. This application is carried out with emphasis in the Portuguese Wheat Breeding Program (1986 - 2000).

It is shown how to adjust a model for environmental indexes with dummy variables associated to years, locations and cultivars. The year coefficients were used to express, for the Portuguese Wheat Breeding Program, year effects.

The results obtained in the studied case clearly indicate the usefulness of JRA. Namely, the technique is useful in deciding which of the cultivars used should be kept for the next year.

Lista de símbolos e siglas

ACR	Análise conjunta de regressão	1
L_2	Índices L_2	1
$\boldsymbol{\theta}_b$	vector do parâmetro de estrutura com b componentes ...	6
$\mathbf{X}(\boldsymbol{\theta}_b)$	matriz do modelo que depende do vector de estrutura $\boldsymbol{\theta}_b$	6
$n \times k$	n linhas e k colunas numa matriz	6
$\boldsymbol{\gamma}_k$	vector dos coeficientes para as regressões lineares com k componentes	6
\mathbf{y}_n	vector de observações da variável dependente com n componentes	6
$E(\mathbf{y}_n)$	valor esperado do vector \mathbf{y}_n	6
$R[\mathbf{X}(\boldsymbol{\theta}_b^o)]$	espaço imagem da matriz $\mathbf{X}(\boldsymbol{\theta}_b^o)$	6
$S(\boldsymbol{\gamma}_k, \boldsymbol{\theta}_b)$	função objectivo a minimizar	6
$\boldsymbol{\theta}_{1,b}$	vector de estrutura inicial com b componentes	6
$S(\boldsymbol{\gamma}_k \boldsymbol{\theta}_{1,b})$	função objectivo a minimizar em ordem a $\boldsymbol{\gamma}_k$ dado $\boldsymbol{\theta}_{1,b}$..	6
$\tilde{\boldsymbol{\gamma}}_{1,k}$	estimador para $\boldsymbol{\gamma}_k$ obtido na primeira iteração	6
\aleph	define a família das matrizes do modelo	7
ς	relação de equivalência entre matrizes do modelo	7
Γ	espaço paramétrico dos vectores de estrutura	7
$\triangleleft \mathbf{v}_b$	mínimo das componentes do vector \mathbf{v}_b	7
$\triangleright \mathbf{v}_b$	máximo das componentes do vector \mathbf{v}_b	7
$\tilde{\boldsymbol{\theta}}_{1,b}$	estimador para $\boldsymbol{\theta}_b$ obtido na primeira iteração	7
Ω^\perp	complemento ortogonal do espaço Ω	7
$\mathbf{y}_{\Omega^\perp}$	projectão ortogonal de \mathbf{y} sobre Ω^\perp	7
$\ \mathbf{a}\ $	norma do vector \mathbf{a}	7
J	número de cultivares utilizados no estudo	8
\mathbf{y}_{j,n_j}	vector dos rendimentos para o j -ésimo cultivar com n_j componentes	8
n	número total de observações no modelo	8
$\mathbf{D}(\mathbf{X}_1, \dots, \mathbf{X}_J)$	matriz diagonal por blocos com blocos principais, $\mathbf{X}_1, \dots, \mathbf{X}_J$	8
$\mathbf{1}_{n_j}$	vector com n_j componentes iguais a um	8

n_j	número de observações para o j -ésimo cultivar 8
\mathbf{x}_{j,n_j}	vector com n_j componentes, sendo a i -ésima componente o índice ambiental do bloco onde se obteve a i -ésima produção do j -ésimo cultivar 8
γ_{2J}	vector dos coeficientes para as regressões 8
$p_{i,j}$	peso atribuído ao j -ésimo cultivar no i -ésimo bloco 8
b	número de pares (local, ano) ou índices ambientais 8
x_i	índice ambiental do i -ésimo bloco, par (local,ano) 8
\mathbf{x}_b	vector de estrutura cujas componentes são os b índices ambientais 8
$y_{i,j}$	rendimento do j -ésimo cultivar no i -ésimo bloco 8
$x_{i,0}$	valor inicial para o índice ambiental do i -ésimo bloco 9
(α_j, β_j)	ordenada na origem e coeficiente angular da recta de regressão a ajustar para o j -ésimo cultivar 9
$S_1(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$	função objectivo a minimizar em ordem a $\boldsymbol{\alpha}_J$ e $\boldsymbol{\beta}_J$ na primeira iteração do algoritmo Zig-Zag 9
$(\tilde{\alpha}_{j,1}, \tilde{\beta}_{j,1})$	estimadores dos mínimos quadrados para (α_j, β_j) obtidos na primeira iteração 9
$S_1(\mathbf{x}_b)$	função a minimizar em ordem a \mathbf{x}_b na primeira iteração do algoritmo Zig-Zag 9
$\min[x_{1,o}, \dots, x_{b,o}]$	representa o mínimo dos elementos do conjunto [] 9
$\max[x_{1,o}, \dots, x_{b,o}]$	representa o máximo dos elementos do conjunto [] 9
$\mathbb{R}_b \times \mathbb{R}_J \times \mathbb{R}_J$	espaço obtido pelo produto cartesiano 9
$a \varphi b$	a é φ equivalente a b 9
$R(\mathbf{1}_b)$	espaço imagem das matrizes coluna com todas as componentes iguais a 1 10
$\text{Int}(a)$	parte inteira de a 13
$\text{grad}(S_1)$	gradiente da função S_1 14
$\text{Hess}(S_1)$	matriz hessiana de S_1 15
$U_{j,l}(u, v)$	função objectivo associada ao j -ésimo cultivar e l -ésimo ano 19
\tilde{S}	Soma das somas dos quadrados dos resíduos das regressões correspondentes à última iteração do algoritmo Zig-Zag 20
$\mathbf{x} \sim N[\mathbf{a}, \mathbf{B}]$	o vector aleatório \mathbf{x} tem distribuição Normal com valor médio \mathbf{a} e matriz de covariância \mathbf{B} 20
g	número total de graus de liberdade associados a \tilde{S} 20
b_j	número de blocos em que o cultivar j está presente 20
h	número de anos considerados no plano de melhoramento 21
$(\tilde{\alpha}_{j,l}, \tilde{\beta}_{j,l})$	coeficientes de regressão ajustados para o j -ésimo cultivar no l -ésimo ano 21

$H_{0,j,j'}(1, v_o)$	hipótese nula relativa aos j -ésimo e j' -ésimo cultivares e à função $U(1, v_o)$	23
$t_{j,j'}(1, v_o)$	valor da estatística do teste t para comparação dos cultivares j e j'	23
$F_{1-\alpha, v_1, v_2}$	valor crítico correspondente à distribuição F de Snedcor central, com v_1 e v_2 graus de liberdade, para a o nível de significância α	23
D_j	conjunto dos índices dos anos em que o j -ésimo cultivar foi utilizado	24
\mathbf{A}'	transposta da matriz \mathbf{A}	24
$\mathbf{diag}(r_1, \dots, r_m)$	matriz diagonal com elementos principais r_1, \dots, r_m	24
$\mathit{Car}(\mathbf{A})$	característica da matriz \mathbf{A}	24
s_{u_i}	soma dos quadrados dos desvios para a média dos índices ambientais ajustados no ano com índice u_i	25
$\mathbf{A} \otimes \mathbf{B}$	produto de Kronecker de \mathbf{A} por \mathbf{B}	27
$\mathit{Var}(U)$	variância da variável U	27
$\mathit{Cov}(\tilde{\mathbf{U}})$	matriz de covariância do vector $\tilde{\mathbf{U}}$	27
C_i	conjunto dos índices dos cultivares presentes no i -ésimo bloco	28
$l(\gamma_{2J}, \boldsymbol{\theta}_b, \sigma^2)$	logaritmo da função de máxima verosimilhança	29
$\hat{\sigma}^2$	estimador de máxima verosimilhança para σ^2	29
$(\hat{\gamma}_{2J}, \hat{\boldsymbol{\theta}}_b)$	estimadores de máxima verosimilhança para o par $(\gamma_{2J}, \boldsymbol{\theta}_b)$	29
w	espaço paramétrico associado a H_o	29
Ω	espaço paramétrico total	29
$L(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J, \boldsymbol{\theta}_b, \sigma^2)$	função de máxima verosimilhança	29
$x_{g, 1-\alpha}$	quantil para a probabilidade $1 - \alpha$ da distribuição do qui-quadrado com g graus de liberdade	36
\approx	aproximadamente	36
$\mathit{prob}(\cdot)$	probabilidade de (\cdot)	36
$\theta_{j,j'}$	ponto de intersecção entre as rectas de regressão ajustadas com índices j e j'	40
\mathbf{V}_b	vector dos índices ambientais no modelo para os índices ambientais	80
p_i	número de colunas da matriz do modelo associadas à i -ésima variável explicativa	80
\mathbf{X}	matriz do modelo para os índices ambientais	81
$\boldsymbol{\delta}_h$	vector dos coeficientes associados à matriz do modelo ..	81
\mathbf{e}_b	vector dos erros no modelo para os índices ambientais ..	81
g	número de valores próprios não nulos da matriz $\mathbf{X}'\mathbf{X}$..	81
$\tilde{\gamma}_g$	estimador de mínimos quadrados usando componentes principais	82

S_{cp}	soma dos quadrados dos resíduos para o modelo estimado usando componentes principais82
$tr(\mathbf{A})$	traço da matriz \mathbf{A} 83
$\tilde{\alpha}$	ordenada na origem da recta de regressão linear ajustada 87
$\tilde{\beta}$	coeficiente angular da recta de regressão linear ajustada 87

Ao longo deste trabalho os vectores e as matrizes estão representados a negrito. O número de componentes de um vector é o valor assumido pelo índice inferior do mesmo. Quando um vector tem dois índices inferiores, o índice à esquerda da vírgula identifica o vector e o índice à direita da vírgula representa o número de componentes do mesmo. Quando uma componente de um vector tem dois índices inferiores, o que está à esquerda da vírgula posiciona a componente no vector em que se insere e o índice à direita da vírgula refere-se ao vector.

Conteúdo

Lista de símbolos e siglas	IX
1 Introdução	1
2 Análise conjunta de Regressões	5
2.1 Nota Histórica	5
2.2 Ajustamento	6
2.2.1 Mínimos Quadrados Estruturados	6
2.2.2 Análise conjunta de regressão e Índices L_2	7
2.2.3 Redes Fraccionalmente completas	17
2.3 Inferência	18
2.3.1 Algumas considerações preliminares	18
2.3.2 Funções Objectivo Lineares	19
2.3.3 Hipóteses sobre um factor	23
2.3.4 Hipóteses sobre dois factores com interacção	25
2.3.5 Testes de quociente de Verosimilhanças	28
2.3.6 Regiões de confiança	36
2.4 Classificação e Selecção de Cultivares	37
2.4.1 Classificação	37
2.4.2 Selecção pelo método do Contorno Superior	39
3 Plano de Melhoramento de Trigo Mole	47
3.1 Considerações Prévias	47
3.2 Índices Ambientais	51

3.3	Estudo da estabilidade dos cultivares	54
3.4	Classificação dos cultivares na Opção <i>I</i>	57
3.5	Análise do Contorno Superior na opção <i>I</i>	60
3.6	Classificação dos cultivares na Opção <i>II</i>	64
3.7	Análise do contorno superior na Opção <i>II</i>	65
3.8	Seleção das funções objectivo	67
3.9	Máximos das funções objectivo	69
3.10	Análise dos grupos de Cultivares	70
4	Índices Ambientais	79
4.1	Considerações Prévias	79
4.2	Modelo	79
4.3	Ajustamento	84
4.4	Funções objectivo	86
5	Conclusões e Trabalhos Futuros	89
A		91
Bibliografia		106

Lista de Tabelas

2.1	Grelha de classificação dos cultivares	38
3.1	Lista dos cultivares	48
3.2	Locais e anos onde se realizaram as experiências	48
3.3	Presença e ausência dos cultivares nos vários anos	49
3.4	Médias dos índices ambientais por bloco	52
3.5	Correlações lineares e de Spearman entre pares de índices	53
3.6	Coefficientes angulares ajustados nas Opções <i>I</i> e <i>II</i>	54
3.7	Soma dos quadrados dos Resíduos para a média dos índices ambientais obtidos na Opção <i>II</i>	55
3.8	Resultados dos testes F	56
3.9	Valores estimados para α , β e α'	58
3.10	Classificação dos cultivares na Opção <i>I</i>	59
3.11	Resultados do ajustamento	60
3.12	Resultados do teste t - Cultivares dominantes e dominados	63
3.13	Resultados dos métodos de comparação múltipla de Bonferroni e de Scheffé ao nível de 5%	64
3.14	Classificação dos cultivares na Opção <i>II</i>	65
3.15	Composição dos contornos superiores ano a ano	66
3.16	Funções objectivo utilizadas	68
3.17	Máximos das funções objectivo	69
3.18	Agrupamento dos cultivares segundo a presença nos anos	70
3.19	Valores estimados para a função f_1 relativos aos cultivares do Grupo quatro	71
3.20	Resultados dos Testes F para efeitos e interacções	73

3.21	Método de Scheffé e Bonferroni para os cultivares do Grupo 3	75
3.22	Método de Scheffé e Bonferroni para os cultivares do Grupo 4	75
3.23	Método de Scheffé para os anos do Grupo 4	75
3.24	Método de Scheffé para os anos do Grupo 8	75
3.25	Método de Scheffé e Bonferroni para os cultivares do Grupo 14	76
3.26	Método de Scheffé e Bonferroni para os cultivares do Grupo 16	76
3.27	Método de Scheffé para os anos do Grupo 18	76
3.28	Método de Scheffé para os anos do Grupo 19	76
4.1	Coeficientes ajustados e testes t na Opção <i>I</i>	85
4.2	Coeficientes ajustados e testes t na Opção <i>II</i>	86
4.3	Coeficientes ajustados para os máximos das funções objectivo	87

Lista de Figuras

2.1	Ponto de Intersecção	41
2.2	Rectas Paralelas	41
2.3	Traçado do contorno superior quando $\bar{\theta}_1 \leq \theta_{min}$	42
2.4	Traçado do contorno superior quando $\theta_{min} < \bar{\theta}_1 < \theta_{max}$	43
2.5	Traçado do contorno superior quando $\bar{\theta}_1 \geq \theta_{max}$	43
2.6	Contorno superior com dois cultivares dominantes	45
2.7	Comparação entre cultivares dominantes e não dominantes	45
3.1	Contorno Superior na Opção <i>I</i>	61
A.1	Contorno Superior no ano de 1986	96
A.2	Contorno Superior no ano de 1987	97
A.3	Contorno Superior no ano de 1988	98
A.4	Contorno Superior no ano de 1989	99
A.5	Contorno Superior no ano de 1990	100
A.6	Contorno Superior no ano de 1991	101
A.7	Contorno Superior no ano de 1992	102
A.8	Contorno Superior no ano de 1995	103
A.9	Contorno Superior no ano de 1997	104
A.10	Contorno Superior no ano de 1999	105

Capítulo 1

Introdução

Um dos aspectos importantes do desenvolvimento duma técnica, corresponde ao alargamento do seu domínio de aplicação.

Na sua formulação usual a Análise Conjunta de Regressões, ACR, enquadra-se no conjunto das técnicas utilizadas, ver Aastveit e Mejza (1992), para o estudo das interacções (genótipos x ambiente). Mais concretamente, pertence ao grupo das técnicas baseadas no ajustamento de regressões. Estas regressões são lineares, uma por cultivar, tendo como variável dependente o rendimento. A variável controlada, o chamado índice ambiental, mede, para cada par (local, ano), a respectiva capacidade produtiva.

Desde a origem da sua aplicação e durante alguns anos a ACR era aplicada a redes de ensaios em blocos casualizados. Seguindo-se Gusmão (1985) e (1986 a) os índices ambientais correspondentes aos diferentes blocos eram medidos pelos respectivos rendimentos médios. Posteriormente, os ensaios passaram a ser sobretudo do tipo α - designs, logo com blocos incompletos, ver Patterson and Williams (1976). Deixou pois, de se poder medir os índices ambientais dos blocos através dos respectivos rendimentos médios. Para ultrapassar esta dificuldade, foram introduzidos os índices ambientais L_2 , ver Mexia et al (1999). Estes índices são ajustados juntamente com os coeficientes das regressões, através da minimização duma função objectivo quadrática. Pereira (2004) desenvolve duma forma sistemática a ACR, utilizando índices ambientais L_2 . No entanto, subjacente ao seu tratamento, estão os pressupostos de que os ensaios foram realizados num número reduzido de anos e que se trabalhou com um conjunto dado de cultivares.

Quando passamos a considerar planos de melhoramento, temos ensaios realizados ao longo dum número considerável de anos e um conjunto aberto de cultivares. Com efeito, ao longo da realização do plano de melhoramento, haverá entrada e saída de cultivares.

O nosso objectivo é mostrar como aplicar a ACR na condução de planos de melhoramento. Dada a singularidade deste desenvolvimento, resolvemos apoiar-mo-nos no estudo dum plano

de melhoramento dado: o do Trigo Mole em Portugal (1986 - 2000). O nosso trabalho, para além do primeiro capítulo correspondente à Introdução e do quinto capítulo correspondente às Conclusões Finais e Trabalhos Futuros, articula-se nos três capítulos:

2. Análise Conjunta de Regressões;
3. Plano de Melhoramento do Trigo Mole;
4. Índices Ambientais.

No primeiro destes capítulos, desenvolvemos os aspectos da ACR de que necessitamos para a discussão de planos de melhoramento. Começamos por estudar as questões relativas ao ajustamento. Observe-se que, quando se utilizam índices ambientais L_2 , se estimam não só os coeficientes das regressões como também os índices ambientais. Classicamente, quando se ajustava uma regressão linear, a matriz do modelo era suposta conhecida. Neste enquadramento generalizado a que chamamos mínimos quadrados estruturados a matriz do modelo é apenas conhecida a menos dum vector de estrutura que, no caso da ACR, é o vector dos índices ambientais. Consideramos em particular o caso de redes de ensaios que se fraccionam em sub-redes, constituídas por ensaios com blocos, contendo todos os cultivares presentes na sub-rede. Havendo uma ordem entre sub-redes e cultivares comuns a redes adjacentes, a análise destes casos apresenta, como veremos, propriedades interessantes. Particularmente, trataremos o caso em que cada sub-rede corresponde a um ano, dando assim lugar a dois tipos de ajustamentos. O ajustamento global em que se utilizam os dados de todos os anos disponíveis, o qual designamos por Opção *I* e o conjunto dos ajustamentos correspondentes a cada ano que designamos por Opção *II*. Segue-se o estudo de questões de inferência estatística. Procura-se fundamentalmente desenvolver técnicas, para avaliar a estabilidade dos comportamentos dos cultivares e para comparar cultivares. Utilizamos funções lineares do tipo $U_{j,l} = \alpha_{j,l} + \beta_{j,l}x$, onde os índices inferiores j e l representam o cultivar e o ano em análise respectivamente. O valor de x pertence ao intervalo de variação dos índices ambientais ajustados. Este valor deve ser escolhido de acordo com os objectivos dos melhoradores. Com efeito, ao utilizar-se uma destas funções estão-se a comparar os valores ajustados para os cultivares e para esse valor do índice ambiental.

Dada a sua importância teórica, seguindo Pereira (2004), mostramos como, se se admitir a normalidade dos vectores dos rendimentos, a introdução dos índices ambientais L_2 faz com que os estimadores que se obtêm para os índices ambientais e coeficientes de regressão, sejam estimadores de máxima verosimilhança. Completamos ainda a bateria de testes de quociente de verosimilhanças dada por Pereira (2004). A finalizar este capítulo, apresentamos um parágrafo sobre classificação e selecção de cultivares. O método do contorno superior, ver Mexia et al (1997), é aqui utilizado como instrumento para a selecção de cultivares.

No terceiro capítulo, tentamos mostrar uma aplicação da ACR a planos de melhoramento, em particular o do Trigo Mole em Portugal (1986 - 2000). O primeiro problema que se põe é o da verificação da estabilidade dos resultados. Este problema é analisado por duas etapas. Na primeira, procura-se relacionar os índices ambientais ajustados globalmente com os ajustados ano a ano. Verifica-se, para o caso analisado, um alto grau de compatibilidade entre os dois conjuntos de índices ambientais. Na segunda etapa, estuda-se a estabilidade dos cultivares utilizados em vários anos. Designam-se por β -estável os cultivares para os quais não se rejeita a hipótese de igualdade dos coeficientes angulares das rectas de regressão ajustadas, em vários anos, para cada um desses cultivares. Na aplicação prática que vamos apresentar, apenas três cultivares em 22 não se revelam β -estáveis.

Verificada a existência de estabilidade dos resultados, podemos aceitar a utilização da ACR e prosseguir a nossa análise. Passamos então à classificação e selecção dos cultivares, o que é feito para cada uma das duas opções. No que respeita à classificação, os cultivares cuja recta de regressão ajustada tem coeficiente angular β superior[inferior] a um são intensivos[extensivos], interessando verificar se β difere significativamente de um. Relativamente à selecção, utiliza-se o método do contorno superior definido pelas regressões ajustadas. Este contorno superior é uma poligonal convexa. Os cultivares cujas rectas de regressão ajustadas integram o contorno superior, são cultivares dominantes. A cada um destes corresponde um intervalo de dominância. Os restantes cultivares, devem ser comparados com os cultivares dominantes, para ver se são significativamente dominados. Dado pretender-se cultivares que aproveitam adequadamente boas condições de fertilidade, são sobretudo importantes os cultivares dominantes correspondentes aos intervalos de dominância à direita. Deve salientar-se que, no caso do melhoramento do trigo mole, para a opção *I*, os cultivares dominantes à direita permaneceram no programa até ao final.

No estudo dos contornos superiores e da classificação dos cultivares ano a ano verificamos a presença de algumas situações difíceis de explicar: cultivares que num ano eram dominantes à direita, β -estável e classificados como intensivos, por vezes não foram utilizados no ano seguinte. Este é um ponto muito interessante, pois aponta para a utilidade da ACR como instrumento na tomada de decisão, relativamente à permanência de cultivares no plano de melhoramento.

Com o objectivo de avaliar o progresso do plano de melhoramento, estudamos em seguida, a evolução dos valores máximos tomados pelas funções objectivo $U_{j,l}$. No entanto, essa evolução revelou-se difícil de interpretar directamente, sendo por isso sujeita a uma análise posterior.

Um último ponto que nos interessa analisar neste capítulo, está relacionado com a existência de grupos de cultivares presentes nos mesmos anos. Seria de supor que os cultivares que

constituem esses grupos tivessem tido um comportamento homogéneo. Foi no entanto possível, detectar frequentes violações desta regra, tendo sido nalguns anos mantidos cultivares a eliminar e eliminados, noutros anos, cultivares dominantes. Verificou-se assim mais uma vez, o interesse da ACR como técnica de acompanhamento dos planos de melhoramento. Resumindo, neste capítulo procurámos, revendo-nos no Plano de Melhoramento do Trigo Mole, mostrar como é possível:

- validar a utilização da ACR na análise dos resultados de planos de melhoramento;
- utilizar, através da análise dos resultados de cada ano, a ACR como técnica de acompanhamento de planos de melhoramento.

O quarto capítulo surge com a tentativa de se poder interpretar melhor os valores máximos tomados pelas funções objectivo. Com efeito, estes apresentam oscilações irregulares de ano para ano, muito possivelmente atribuíveis a factores de carácter climático. Para testar esta conjectura, ajusta-se um modelo linear para os índices ambientais. Neste modelo consideram-se variáveis dummy correspondentes a:

- anos;
- locais;
- grupos de cultivares.

Os grupos de cultivares são os mesmos que atrás tinham sido considerados. Os coeficientes das variáveis dummy associados a anos são utilizados para exprimir as capacidades produtivas dos vários anos. É assim possível quantificar a influência do factor ano nos máximos das funções objectivo.

Para terminar a fase da introdução, resta-nos ainda dizer que os programas informáticos utilizados para a execução de todos os cálculos inerentes à aplicação que apresentamos neste trabalho, foram desenvolvidos no Maple. Este programa é bastante funcional, uma vez que permite em simultâneo programar e utilizar funções de cálculo já existentes.

Capítulo 2

Análise conjunta de Regressões

2.1 Nota Histórica

A iniciar este trabalho vamos referir brevemente a história da Análise Conjunta de Regressões Lineares. Mooers (1921) and Yates & Cochran (1938), foram os pioneiros no desenvolvimento desta técnica. Mais tarde, nos anos sessenta, Finlay & Wilkinson (1963) e Eberhart & Russel (1966), deram o seu grande contributo para aplicação da técnica à comparação de cultivares sujeitos a um vasto número de níveis de fertilidade. Como contributo português, Gusmão (1985), (1986a) e (1986b), mostrou que a precisão dos ajustamentos melhorava grandemente se se considerassem índices ambientais distintos para os diferentes blocos, em vez dum único índice por ensaio.

Na década de oitenta, tendo como argumento o facto de, na ACR não se considerarem variáveis ambientais específicas, Westcott (1986) e Lin et al. (1986), criticaram o uso desta técnica em oposição a Becker & Leon (1988) que consideram ser uma crítica pouco importante. Em Portugal temos vários autores que se debruçaram sobre este tema, são eles: Gusmão (1986b), (1990), Gusmão et al. (1991a), (1991b), (1992a), (1992b), (1993a), (1993b), (1995), Ferreira et al. (1990), Maças et al. (1990), Mexia et al. (1991) e Dias (2000). Seguindo-se Gusmão (1985) e (1986a), a ACR era aplicada a redes de ensaios em blocos casualizados, sendo os índices ambientais correspondentes aos diferentes blocos medidos pelos respectivos rendimentos médios. Este procedimento só se podia aplicar quando se tinham blocos completos. Assim, Mexia et al. (1999), com a introdução dos índices ambientais L_2 , alargaram o campo de utilização da ACR ao caso incompleto.

Digby (1979) e Ng & Grunwald (1997) desenvolveram também métodos iterativos para o ajustamento conjunto de regressões.

Refira-se ainda, a dissertação de doutoramento de Pereira (2004), em que a ACR utilizando índices ambientais L_2 , é desenvolvida numa forma sistemática. Aliás, o presente trabalho

representa uma extensão dessa dissertação, a planos de melhoramento.

2.2 Ajustamento

2.2.1 Mínimos Quadrados Estruturados

Ao formular-se um modelo utiliza-se a informação disponível para exprimir, sob a forma de equações, as relações entre as variáveis intervenientes: controladas e dependentes. No caso duma única variável dependente, relacionada linearmente com as variáveis explicativas, os valores destas, correspondentes a uma observação da variável dependente, constituem uma linha da matriz do modelo. Neste caso, a matriz do modelo é conhecida, desde que se conheçam os valores das variáveis explicativas, o que nem sempre se verifica. Por exemplo, no caso da ACR, os índices ambientais que medem a produtividade de cada par (local, ano) terão de ser estimados. Os mínimos quadrados estruturados, surgiram como uma técnica para estimar, não só os coeficientes das regressões, como também os valores das variáveis explicativas. Admite-se para isso, que a matriz do modelo é conhecida, a menos de um parâmetro de estrutura θ_b . As matrizes do modelo serão representadas por $\mathbf{X}(\theta_b)$, sendo do tipo de $n \times k$ com $b < n$ e $k < n$, e o vector dos coeficientes por γ_k . Sendo \mathbf{y}_n o vector das observações, admite-se que o vector médio $E(\mathbf{y}_n)$, pertence ao espaço imagem $\Omega^o = R[\mathbf{X}(\theta_b^o)]$, com θ_b^o o verdadeiro valor do parâmetro de estrutura. Obtém-se assim uma família parametrizada de modelos. Para estimarmos conjuntamente as componentes dos vectores γ_k e θ_b , minimiza-se

$$S(\gamma_k, \theta_b) = \|\mathbf{y}_n - \mathbf{X}(\theta_b)\gamma_k\|^2, \quad (2.2.1)$$

admitindo-se para isso que as derivadas parciais desta função objectivo são contínuas. Para realizar o ajustamento, pode-se utilizar um algoritmo designado Zig-Zag, ver Mexia et al (1999), por nele se minimizar alternadamente em ordem às componentes do vector de coeficientes e do vector de estrutura. Partindo dum vector inicial, $\theta_{1,b}$, minimiza-se

$$S(\gamma_k | \theta_{1,b}) = \|\mathbf{y}_n - \mathbf{X}(\theta_{1,b})\gamma_k\|^2 \quad (2.2.2)$$

em ordem às componentes do vector γ_k . Sendo $\tilde{\gamma}_{1,k}$ o mínimo obtido, minimiza-se em seguida

$$S(\theta_b | \tilde{\gamma}_{1,k}) = \|\mathbf{y}_n - \mathbf{X}(\theta_b)\tilde{\gamma}_{1,k}\|^2 \quad (2.2.3)$$

agora em ordem às componentes do vector θ_b . Como adiante na subsecção 2.2.2 iremos particularizar a aplicação desta abordagem ao estudo da ACR, interessa-nos verificar a possibilidade de se proceder a um ajustamento de escala para os vectores de estrutura. Para

esse efeito, sendo \aleph a família das matrizes do modelo, pode-se definir em \aleph uma relação de equivalência ς , escrevendo

$$\mathbf{X}(\dot{\boldsymbol{\theta}}_b) \varsigma \mathbf{X}(\ddot{\boldsymbol{\theta}}_b) \quad (2.2.4)$$

quando se tem

$$R[\mathbf{X}(\dot{\boldsymbol{\theta}}_b)] = R[\mathbf{X}(\ddot{\boldsymbol{\theta}}_b)]. \quad (2.2.5)$$

Em geral, pode admitir-se que os vectores de estrutura identificam as matrizes do modelo, tendo-se $\mathbf{X}(\dot{\boldsymbol{\theta}}_b) = \mathbf{X}(\ddot{\boldsymbol{\theta}}_b)$ se e só se $\dot{\boldsymbol{\theta}}_b = \ddot{\boldsymbol{\theta}}_b$. Sendo Γ o espaço paramétrico dos vectores de estrutura, poremos $\dot{\boldsymbol{\theta}}_b \varsigma \ddot{\boldsymbol{\theta}}_b$, quando (2.2.5) se verificar. O caso que nos interessa, como veremos adiante, é aquele em que ς é uma relação de homótopia, isto é quando $\forall d$,

$$(c\boldsymbol{\theta}_b + d\mathbf{1}_b) \varsigma \boldsymbol{\theta}_b \quad (2.2.6)$$

sempre que $c \neq 0$. Diremos então que o modelo é homótopico. A expressão anterior subentende que, dados $\boldsymbol{\theta}_b \in \Gamma$ e $c \neq 0$, $c\boldsymbol{\theta}_b + d\mathbf{1}_b \in \Gamma$. Representando por $\triangleleft \mathbf{v}_b$ e $\triangleright \mathbf{v}_b$ respectivamente o mínimo e o máximo das componentes de \mathbf{v}_b , podemos, no caso dos modelos homotópicos, substituir o mínimo $\tilde{\boldsymbol{\theta}}_{1,b}$, obtido pela minimização de $S(\boldsymbol{\theta}_b | \tilde{\boldsymbol{\gamma}}_{1,k})$, por $\tilde{\boldsymbol{\theta}}_{1,b}^*$ tal que

$$\left\{ \begin{array}{l} \tilde{\boldsymbol{\theta}}_{1,b}^* \varsigma \tilde{\boldsymbol{\theta}}_{1,b} \\ \triangleleft \tilde{\boldsymbol{\theta}}_{1,b}^* = \triangleleft \tilde{\boldsymbol{\theta}}_{1,b} \\ \triangleright \tilde{\boldsymbol{\theta}}_{1,b}^* = \triangleright \tilde{\boldsymbol{\theta}}_{1,b}. \end{array} \right. \quad (2.2.7)$$

Como, com $\Omega = R[\mathbf{X}(\boldsymbol{\theta}_b)]$ e Ω^\perp o complemento ortogonal de Ω se tem,

$$\text{Min}_{\boldsymbol{\gamma}_k} S(\boldsymbol{\gamma}_k | \boldsymbol{\theta}_b) = \|\mathbf{y}_{\Omega^\perp}\|^2, \quad (2.2.8)$$

virá

$$\text{Min}_{\boldsymbol{\gamma}_k} S(\boldsymbol{\gamma}_k | \tilde{\boldsymbol{\theta}}_{1,b}^*) = \text{Min}_{\boldsymbol{\gamma}_k} S(\boldsymbol{\gamma}_k | \tilde{\boldsymbol{\theta}}_{1,b}). \quad (2.2.9)$$

Pode agora iniciar-se um segundo ciclo com $\boldsymbol{\theta}_b = \tilde{\boldsymbol{\theta}}_{1,b}^*$ e repetir as minimizações até que o valor tomado pela função objectivo estabilize. Dado proceder-se por minimizações sucessivas, os mínimos que se vão obtendo no final de cada ciclo, constituem uma sucessão decrescente, limitada inferiormente por zero, logo convergente.

2.2.2 Análise conjunta de regressão e Índices L_2

A ACR, é como referimos uma técnica para comparar cultivares e para estudar a interacção entre genótipos e ambientes. Os ambientes serão os pares (local, ano). Se os ensaios

estiverem divididos em blocos, Gusmão (1985) mostrou que a precisão dos ajustamentos melhora grandemente ao considerarem-se índices ambientais distintos para os diferentes blocos, em vez de um único índice por ensaio.

Tendo-se J cultivares para comparar, a ACR assenta no ajustamento de regressões lineares dos rendimentos dos J cultivares com que se trabalha, nos índices ambientais. Como estes últimos são desconhecidos, também terão de ser estimados, o que faremos através do algoritmo Zig-Zag.

O nosso objectivo neste trabalho, é tratar o caso em que o conjunto de cultivares presente em cada ano vai mudando, e os locais utilizados também não são sempre os mesmos. Assim, para utilizarmos uma formulação do tipo da utilizada na subsecção 2.2.1., sendo \mathbf{y}_{j,n_j} o vector composto pelos n_j rendimentos para o cultivar j , sobrepõem-se os vectores $\mathbf{y}_{1,n_1}, \dots, \mathbf{y}_{J,n_J}$ dos rendimentos para os J cultivares com que se trabalha para se obter um único vector \mathbf{y}_n com $n = n_1 + \dots + n_J$ componentes. Uma vez que não estamos no caso completo, ter-se-á uma matriz do modelo diagonal por blocos,

$$\mathbf{X} = \mathbf{D}(\mathbf{X}_1, \dots, \mathbf{X}_J) \quad (2.2.10)$$

com blocos principais

$$\mathbf{X}_j = [\mathbf{1}_{n_j}, \mathbf{x}_{j,n_j}], j = 1, \dots, J \quad (2.2.11)$$

onde as componentes do vector coluna \mathbf{x}_{j,n_j} correspondentes ao j -ésimo cultivar, serão os índices ambientais dos blocos em que o mesmo está presente. O vector dos coeficientes para as regressões será

$$\boldsymbol{\gamma}_{2J} = [\alpha_1, \beta_1, \dots, \alpha_J, \beta_J]'. \quad (2.2.12)$$

Podemos aligeirar o tratamento, ver Dias (2000), através da introdução de pesos $p_{i,j}$, que têm respectivamente o valor 1[0] quando o j -ésimo cultivar está presente[ausente] no i -ésimo bloco. Sendo b o número de pares (local, ano), teremos b índices ambientais, x_1, \dots, x_b ,¹ que serão as componentes do vector de estrutura \mathbf{x}_b e que também terão de ser ajustados, pelo que iremos minimizar

$$S(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J, \mathbf{x}_b) = \sum_{j=1}^J \left[\sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2 \right] \quad (2.2.13)$$

onde $y_{i,j}$ é o rendimento do j -ésimo cultivar no i -ésimo bloco se $p_{i,j} = 1$ e qualquer valor se $p_{i,j} = 0$. Como referimos, há que começar por escolher valores iniciais para as componentes do vector de estrutura \mathbf{x}_b . Quando se têm ensaios com blocos completos, tomam-se os rendimentos médios dos blocos como valores iniciais dos índices ambientais. Se o delineamento for do tipo α , ver Patterson and Williams (1976), os blocos estão agrupados em super-blocos,

¹Como referimos, os vários blocos que existam serão tratados como locais distintos

estando cada cultivar representado α vezes. Nesse caso a cada bloco atribui-se, como valor inicial do respectivo índice ambiental, o rendimento médio para o super-bloco em que está integrado. Observe-se que, ver Mexia e Pinto (2003), em certos casos reais em que não se dispõe de blocos completos ou mesmo de super-blocos como os referidos, o algoritmo Zig-Zag converge com um número reduzido de iterações, desde que se tomem os rendimentos médios dos blocos como os valores iniciais para os respectivos índices ambientais. Tomaremos esta opção e sendo $x_{i,o}, i = 1, \dots, b$, os referidos valores iniciais, começamos por minimizar,

$$S_1(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \left[\sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_{i,o})^2 \right] \quad (2.2.14)$$

em ordem aos vectores $\boldsymbol{\alpha}_J = (\alpha_1, \dots, \alpha_J)$ e $\boldsymbol{\beta}_J = (\beta_1, \dots, \beta_J)$. Vê-se facilmente que o mínimo é atingido pelos estimadores usuais dos mínimos quadrados. Representemos os mesmos por $(\tilde{\alpha}_{j,1}, \tilde{\beta}_{j,1}), j = 1, \dots, J$, onde o índice inferior 1 indica que se trata da primeira iteração. Em seguida, haverá que minimizar para o vector de estrutura, como se estabeleceu em (2.2.3). A função objectivo $S_1(\mathbf{x}_b)$ será

$$S_1(\mathbf{x}_b) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \tilde{\alpha}_{j,1} - \tilde{\beta}_{j,1} x_{i,1})^2 = \sum_{i=1}^b h_{i,1}(x_{i,1}) \quad (2.2.15)$$

com

$$h_{i,1}(x) = \sum_{j=1}^J p_{i,j} (y_{i,j} - \tilde{\alpha}_{j,1} - \tilde{\beta}_{j,1} x)^2, i = 1, \dots, b \quad (2.2.16)$$

sendo fácil de ver que se pode minimizar separadamente cada uma destas funções obtendo-se,

$$\tilde{x}_{i,1} = \frac{\sum_{j=1}^J p_{i,j} \tilde{\beta}_{j,1} y_{i,j} - \sum_{j=1}^J p_{i,j} \tilde{\alpha}_{j,1} \tilde{\beta}_{j,1}}{\sum_{j=1}^J p_{i,j} \tilde{\beta}_{j,1}^2}, i = 1, \dots, b. \quad (2.2.17)$$

Mostremos agora que o modelo é homotópico para, como referimos em (2.2.6) e (2.2.7), podermos manter invariantes o mínimo e o máximo dos índices ambientais utilizados nas várias iterações. Designemos esse mínimo e esse máximo por $a = \min[x_{1,o}, \dots, x_{b,o}]$ e $b = \max[x_{1,o}, \dots, x_{b,o}]$. Interessa-nos assim estabelecer uma relação de equivalência entre trios $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \in \mathbb{R}_b \times \mathbb{R}_J \times \mathbb{R}_J$, de vectores.

Lema 2.1 *Pondo-se $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \varphi(\dot{\mathbf{x}}_b, \dot{\boldsymbol{\alpha}}_J, \dot{\boldsymbol{\beta}}_J)$, quando existirem $c \neq \emptyset$ e d tais que*

$$\begin{cases} \dot{\mathbf{x}}_b &= c\mathbf{x}_b + d\mathbf{1}_b \\ \dot{\boldsymbol{\alpha}}_J &= \boldsymbol{\alpha}_J - \frac{d}{c}\boldsymbol{\beta}_J \\ \dot{\boldsymbol{\beta}}_J &= \frac{1}{c}\boldsymbol{\beta}_J \end{cases} \quad (2.2.18)$$

estabelece-se uma relação de equivalência entre trios de vectores pertencentes a $\mathbb{R}_b \times \mathbb{R}_J \times \mathbb{R}_J$.

Dem: Tomando-se $c = 1$ e $d = 0$ vem, para todo o trio, $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)\varphi(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$, pelo que φ goza da propriedade reflexiva. Admitamos agora que $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)\varphi(\dot{\mathbf{x}}_b, \dot{\boldsymbol{\alpha}}_J, \dot{\boldsymbol{\beta}}_J)$, existindo pois $c \neq \emptyset$ e d tais que 2.2.18 se verifica, então com $\dot{c} = c^{-1}$ e $\dot{d} = -\frac{d}{c}$ tem-se

$$\begin{cases} \mathbf{x}_b &= \dot{c}\dot{\mathbf{x}}_b + \dot{d}\mathbf{1}_b \\ \boldsymbol{\alpha}_J &= \dot{\boldsymbol{\alpha}}_J - \frac{\dot{d}}{\dot{c}}\dot{\boldsymbol{\beta}}_J \\ \boldsymbol{\beta}_J &= \frac{1}{\dot{c}}\dot{\boldsymbol{\beta}}_J \end{cases} \quad (2.2.19)$$

e consequentemente φ goza da propriedade simétrica. Finalmente, admitamos que $(\mathbf{x}_{i,b}, \boldsymbol{\alpha}_{i,J}, \boldsymbol{\beta}_{i,J})\varphi(\mathbf{x}_{i+1,b}, \boldsymbol{\alpha}_{i+1,J}, \boldsymbol{\beta}_{i+1,J})$, existindo pois $c_i \neq 0$ e d_i , tais que (2.2.18) se verifica para $i = 1, 2$. Agora, com $c = c_1c_2 \neq \emptyset$ e $d = d_1c_2 + d_2$, as expressões (2.2.18) verificam-se para o par $[(\mathbf{x}_{1,b}, \boldsymbol{\alpha}_{1,J}, \boldsymbol{\beta}_{1,J}), (\mathbf{x}_{3,b}, \boldsymbol{\alpha}_{3,J}, \boldsymbol{\beta}_{3,J})]$ o que mostra que φ goza da propriedade transitiva, estando a demonstração completa. ■

Sejam $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\varphi$ a classe de equivalência φ a que $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ pertence, $x_o = \frac{1}{b} \sum_{i=1}^b x_i$ e

$D(\mathbf{x}_b) = \sqrt{\sum_{i=1}^b (x_i - x_o)^2}$. Se o vector \mathbf{x}_b tiver todas as suas componentes iguais, pertence ao espaço imagem $R(\mathbf{1}_b)$ da matriz coluna $\mathbf{1}_b$, e nesse caso facilmente se verifica que $D(\mathbf{x}_b) = 0$.

Estabeleçamos o

Lema 2.2 *Se $\mathbf{x}_b \in R(\mathbf{1}_b)$, $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\varphi$ não terá nenhum trio de vectores em $V = \{(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) : (x_o = 0) \wedge (\|\mathbf{x}_b\| = 1)\}$ no entanto se $\mathbf{x}_b \notin R(\mathbf{1}_b)$, $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\varphi$ conterá dois trios em V .*

Dem: Se $\mathbf{x}_b \in R(\mathbf{1}_b)$ as suas componentes são todas iguais e portanto iguais à média x_o , pelo que se tivermos $x_1 = x_2 = \dots = x_b = x_o = 0$, não se verifica a condição $(\|\mathbf{x}_b\| = 1)$, não havendo assim nenhum trio em $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\varphi \cap V$.

Se pelo contrário $\mathbf{x}_b \notin R(\mathbf{1}_b)$ e se $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)\varphi(\dot{\mathbf{x}}_b, \dot{\boldsymbol{\alpha}}_J, \dot{\boldsymbol{\beta}}_J)$ tem-se : $\dot{x}_o = \frac{1}{b} \sum_{i=1}^b \dot{x}_i = c x_o + d$ e $D(\dot{\mathbf{x}}_b) = |c|D(\mathbf{x}_b)$. Ora $\dot{x}_o = 0$ e $D(\dot{\mathbf{x}}_b) = 1$ se e só se $c = \pm D(\mathbf{x}_b)^{-1}$ e $d = -c x_o$ o que mostra que, quando $\mathbf{x}_b \notin R(\mathbf{1}_b)$, $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\varphi$ contem dois trios em V . ■

Quando, nesta subsecção para simplificar escrevermos S , entenda-se $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$.

Estabeleçamos agora a

Proposição 2.1 *A função objectivo*

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2$$

tem valor constante nas classes de equivalência φ , tomando todos os seus valores, para $\mathbf{x}_b \notin R(\mathbf{1}_b)$, em V . O mínimo absoluto da função objectivo é saturado para φ sendo a união de classes de equivalência φ que intersectam V .

Dem: Como, quando $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)\varphi(\dot{\mathbf{x}}_b, \dot{\boldsymbol{\alpha}}_J, \dot{\boldsymbol{\beta}}_J)$, $y_{i,j} - \alpha_j - \beta_j x_i = y_{i,j} - \dot{\alpha}_j - \dot{\beta}_j \dot{x}_i$ a função $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ é constante nas classes de equivalência φ . Assim, o mínimo absoluto da função objectivo, a existir, terá de ser saturado para φ . Vê-se ainda, atendendo ao lema 2.2, que a função objectivo toma todos os seus valores para $\mathbf{x}_b \notin R(\mathbf{1}_b)$, em V .

Por outro lado, se $\mathbf{x}_b \in R(\mathbf{1}_b)$, com $c = 1$ e $d = -x_o$, teremos $\dot{\mathbf{x}}_b = \mathbf{1}\mathbf{x}_b - x_o\mathbf{1}_b = \mathbf{0}_b$, $\dot{\boldsymbol{\alpha}}_J = \boldsymbol{\alpha}_J - x_o\boldsymbol{\beta}_J$ e $\dot{\boldsymbol{\beta}}_J = \boldsymbol{\beta}_J$.

Dado agora $\ddot{\mathbf{x}}_b \notin R(\mathbf{1}_b)$, teremos

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = S(\mathbf{0}_b, \dot{\boldsymbol{\alpha}}_J, \dot{\boldsymbol{\beta}}_J) = S(\ddot{\mathbf{x}}_b, \dot{\boldsymbol{\alpha}}_J, \mathbf{0}_J) \geq S(\ddot{\mathbf{x}}_b, \tilde{\boldsymbol{\alpha}}_J(\ddot{\mathbf{x}}_b), \tilde{\boldsymbol{\beta}}_J(\ddot{\mathbf{x}}_b))$$

com $\tilde{\boldsymbol{\alpha}}_J(\ddot{\mathbf{x}}_b)$ e $\tilde{\boldsymbol{\beta}}_J(\ddot{\mathbf{x}}_b)$ os estimadores de mínimos quadrados dos vectores de coeficientes.

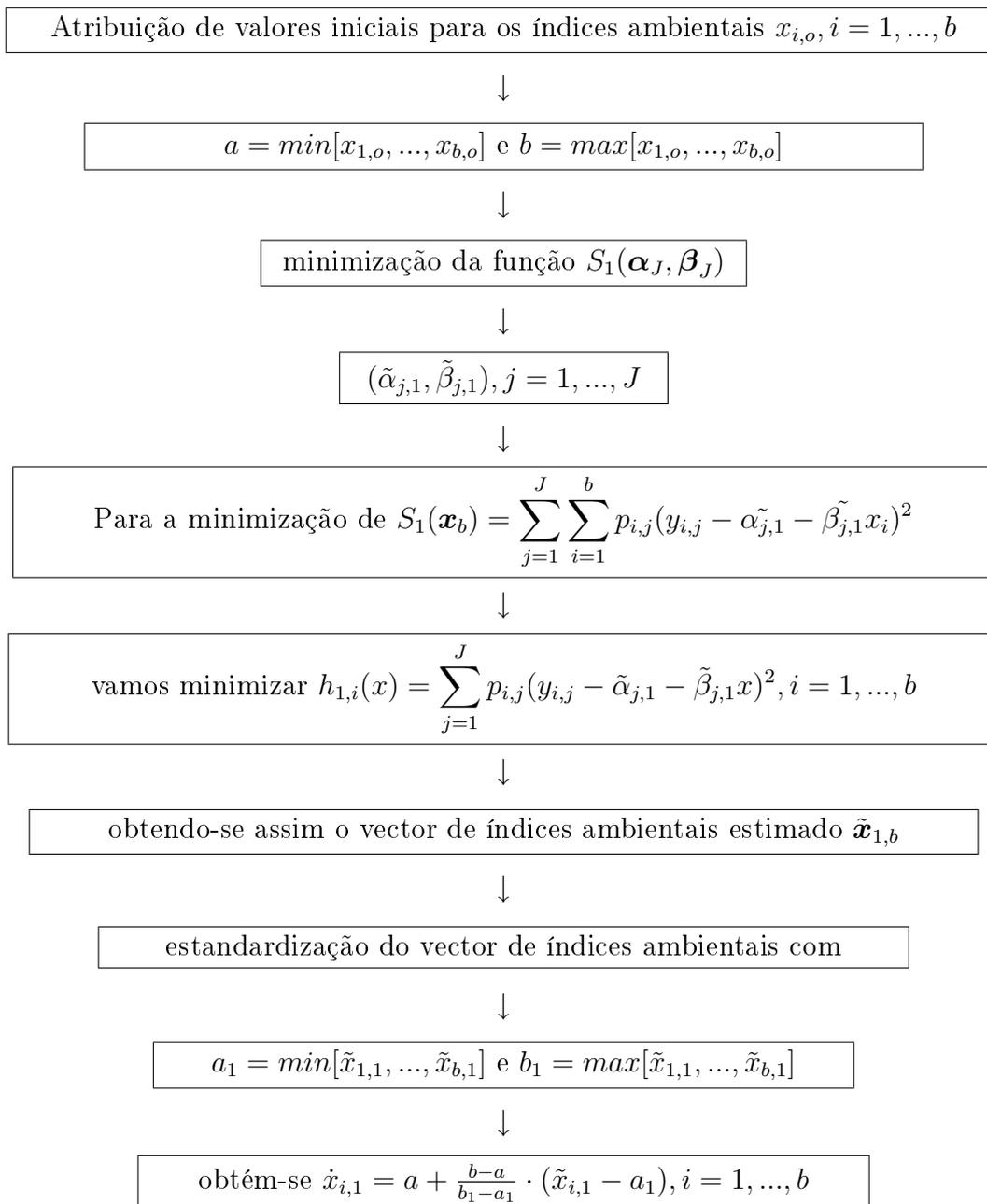
Assim o mínimo da função objectivo, a existir, terá de ser atingido para $\mathbf{x}_b \notin R(\mathbf{1}_b)$.

Resta-nos mostrar que o mínimo absoluto existe. Pondo-se $\tilde{S}(\mathbf{x}_b) = S(\mathbf{x}_b, \tilde{\boldsymbol{\alpha}}_J(\mathbf{x}_b), \tilde{\boldsymbol{\beta}}_J(\mathbf{x}_b))$ tem-se $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \geq \tilde{S}(\mathbf{x}_b)$, pelo que nos bastará mostrar que existe mínimo absoluto para $\tilde{S}(\mathbf{x}_b)$, com $\mathbf{x}_b \notin R(\mathbf{1}_b)$. Ora raciocinando como para obter (2.2.19) obtém-se, com $c \neq 0$, $\tilde{\boldsymbol{\alpha}}_J(c\mathbf{x}_b + d\mathbf{1}_b) = \tilde{\boldsymbol{\alpha}}_J(\mathbf{x}_b) - \frac{d}{c}\tilde{\boldsymbol{\beta}}_J(\mathbf{x}_b)$ e $\tilde{\boldsymbol{\beta}}_J(c\mathbf{x}_b + d\mathbf{1}_b) = \frac{1}{c}\tilde{\boldsymbol{\beta}}_J(\mathbf{x}_b)$, vindo $\tilde{S}(c\mathbf{x}_b + d\mathbf{1}_b) = \tilde{S}(\mathbf{x}_b)$. Se $\mathbf{x}_b \notin R(\mathbf{1}_b)$, podemos escolher c e d por forma a ter-se $c\mathbf{x}_b + d\mathbf{1}_b \in A = \{\mathbf{x}_b : (x_o = 0) \wedge (\|\mathbf{x}_b\| = 1)\}$.

Assim, $\tilde{S}(\mathbf{x}_b)$ tomará todos os seus valores para $\mathbf{x}_b \notin R(\mathbf{1}_b)$, em A . Como A é compacto, vê-se devido ao teorema de Weierstrass, que $\tilde{S}(\mathbf{x}_b)$ atinge um mínimo em A . Este mínimo coincidirá com os mínimos de $\tilde{S}(\mathbf{x}_b)$ e de $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ para $\mathbf{x}_b \notin R(\mathbf{1}_b)$, o qual será ainda, como vimos, o mínimo absoluto de $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$. ■

O seguinte esquema sistematiza a aplicação do algoritmo Zig-Zag.

Esquema 1 - Algoritmo Zig-Zag



Repete-se o procedimento a partir do segundo passo, até que as sucessivas somas das somas dos quadrados dos resíduos, difiram a menos duma constante suficientemente pequena.

Aplicando sucessivamente a minimização alternativa atrás descrita, obtém-se uma sequência decrescente de mínimos, no entanto só existe um mínimo absoluto, como veremos, se a rede de ensaios for conectada. Vejamos em que consiste esta condição. Dados os vários blocos em que as experiências se decompõem, é possível unir todos eles por cadeias de blocos, em que blocos adjacentes, ver por exemplo Kagayema (1984), contêm pelo menos um cultivar em comum.

Com efeito, seguindo Pereira (2004), vamos estabelecer a

Proposição 2.2 *Quando a rede não é conectada, os conjuntos de cultivares presentes nas várias sub-redes conectadas, são disjuntos, existindo mínimos absolutos para a função objectivo que não são equivalentes φ .*

Dem: Se uma rede se divide em duas sub-redes conectadas, caso as mesmas tenham um ou mais cultivares em comum a rede global será conectada, ver Bose(1950), Raghavarao(1971) e Kagayema(1984). Assim, é necessário e suficiente que os conjuntos de cultivares presentes nas duas sub-redes conectadas sejam disjuntos, para que a rede global não seja conectada. Este resultado estende-se directamente ao caso em que, a rede global se decompõe em $L \geq 2$ sub-redes conectadas. Sejam $B_l, l = 1, \dots, L$ os conjuntos de índices dos blocos que integram cada uma das L sub-redes e $C_l, l = 1, \dots, L$ o conjunto dos índices dos cultivares presentes na l -ésima sub-rede com $l = 1, \dots, L$. Tem-se então a função objectivo

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{l=1}^L S_l(\mathbf{x}_{l,b_l}, \boldsymbol{\alpha}_{l,J_l}, \boldsymbol{\beta}_{l,J_l}) = \sum_{l=1}^L \sum_{i \in B_l} \sum_{j \in C_l} p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2. \quad (2.2.20)$$

O valor mínimo tomado por S será a soma de valores mínimos para as sub-funções objectivo correspondentes às sub-redes. Assim, cada mínimo $(\tilde{\mathbf{x}}_b, \tilde{\boldsymbol{\alpha}}_J, \tilde{\boldsymbol{\beta}}_J)$ de S será a reunião de mínimos $(\tilde{\mathbf{x}}_{l,b_l}, \tilde{\boldsymbol{\alpha}}_{l,J_l}, \tilde{\boldsymbol{\beta}}_{l,J_l})$ das $S_l, l = 1, \dots, L$. É agora fácil de ver que se podem ter mínimos de S que não são equivalentes φ , o que estabelece a tese. ■

Recomenda-se assim, a decomposição de redes não conectadas, em sub-redes conectadas, a analisar separadamente. Aliás, a conexão é apenas uma condição necessária, para que os mínimos de S sejam equivalentes φ .

Estabeleça-se a

Proposição 2.3 *Para que os mínimos de S sejam equivalentes φ é necessário e suficiente que,*

(a) *A rede seja conectada;*

(b) $J \leq \text{Int}[\frac{n-b+2}{2}]$

com n o número total de parcelas existentes nos b blocos.

Dem: Atendendo à proposição 2.2 verifica-se que a condição a) é necessária para que os mínimos de S sejam equivalentes φ . No entanto, mesmo que uma rede seja conectada existe ainda a possibilidade de o mínimo de S conter mais do que uma classe de equivalência φ . Assim, uma rede com n parcelas agrupadas em b blocos, com J cultivares para comparar, terá b índices ambientais e $2J$ coeficientes a estimar. Da proposição 2.1 sabe-se que se

podem escolher os mínimos de modo a estarem em V , ou seja, introduzem-se duas condições relativas ao vector dos índices ambientais. Temos assim $b + 2J - 2$ parâmetros a ajustar, logo para que se possa ajustar sem ambiguidade o modelo devemos ter, $n \geq b + 2J - 2$, isto é $J \leq \text{Int}[\frac{n-b+2}{2}]$, como se indica na alínea b), o que completa a demonstração. ■

Dada uma rede conectada que satisfaz a condição $J \leq \text{Int}[\frac{n-b+2}{2}]$, estamos em condições de proceder ao ajustamento utilizando o algoritmo Zig-Zag. Como vimos minimizam-se alternadamente as funções objectivo parciais,

$$\begin{cases} S_1(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J | \mathbf{x}_b) = S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \\ S_2(\mathbf{x}_b | \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \end{cases} \quad (2.2.21)$$

obtidas a partir da função objectivo, fixando o vector dos índices ambientais no primeiro caso e os vectores dos coeficientes no segundo. Minimiza-se a primeira em ordem aos coeficientes e a segunda em ordem aos índices ambientais. No fim de cada ciclo procede-se à standardização dos índices ambientais ajustados. Note-se que em (2.2.14) o índice inferior 1 indica que estamos na primeira iteração do algoritmo Zig-zag, no entanto em (2.2.22) e (2.2.23) os índices inferiores 1 e 2 destinam-se apenas a distinguir as duas funções.

Como se pode ver em Bazarra et al. (1992, pág. 113), funções com matrizes hessiana definidas positivas têm mínimos absolutos únicos. Ora cada uma das funções objectivo parciais têm o gradiente,

$$\text{grad}(S_1) = \begin{bmatrix} \frac{\partial S_1}{\partial \alpha_1} \\ \vdots \\ \frac{\partial S_1}{\partial \alpha_J} \\ \frac{\partial S_1}{\partial \beta_1} \\ \vdots \\ \frac{\partial S_1}{\partial \beta_J} \end{bmatrix} = -2 \begin{bmatrix} \sum_{i=1}^b p_{i,1}(y_{i,1} - \alpha_1 - \beta_1 x_i) \\ \vdots \\ \sum_{i=1}^b p_{i,J}(y_{i,J} - \alpha_J - \beta_J x_i) \\ \sum_{i=1}^b p_{i,1}(y_{i,1} - \alpha_1 - \beta_1 x_i)x_i \\ \vdots \\ \sum_{i=1}^b p_{i,J}(y_{i,J} - \alpha_J - \beta_J x_i)x_i \end{bmatrix} \quad (2.2.22)$$

$$\text{grad}(S_2) = \begin{bmatrix} \frac{\partial S_2}{\partial x_1} \\ \vdots \\ \frac{\partial S_2}{\partial x_b} \end{bmatrix} = -2 \begin{bmatrix} \sum_{j=1}^J p_{1,j}(y_{1,j} - \alpha_j - \beta_j x_1)\beta_j \\ \vdots \\ \sum_{j=1}^J p_{b,j}(y_{b,j} - \alpha_j - \beta_j x_b)\beta_j \end{bmatrix} \quad (2.2.23)$$

e a matriz hessiana,

$$Hess(S_1) = 2 \begin{bmatrix} \frac{\partial^2 S_1}{\partial \alpha_1^2} & \cdots & \frac{\partial^2 S_1}{\partial \alpha_1 \partial \alpha_J} & \frac{\partial^2 S_1}{\partial \alpha_1 \partial \beta_1} & \cdots & \frac{\partial^2 S_1}{\partial \alpha_1 \partial \beta_J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 S_1}{\partial \alpha_J \partial \alpha_1} & \cdots & \frac{\partial^2 S_1}{\partial \alpha_J^2} & \frac{\partial^2 S_1}{\partial \alpha_J \partial \beta_1} & \cdots & \frac{\partial^2 S_1}{\partial \alpha_J \partial \beta_J} \\ \frac{\partial^2 S_1}{\partial \beta_1 \partial \alpha_1} & \cdots & \frac{\partial^2 S_1}{\partial \beta_1 \partial \alpha_J} & \frac{\partial^2 S_1}{\partial \beta_1^2} & \cdots & \frac{\partial^2 S_1}{\partial \beta_1 \partial \beta_J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 S_1}{\partial \beta_J \partial \alpha_1} & \cdots & \frac{\partial^2 S_1}{\partial \beta_J \partial \alpha_J} & \frac{\partial^2 S_1}{\partial \beta_J \partial \beta_1} & \cdots & \frac{\partial^2 S_1}{\partial \beta_J^2} \end{bmatrix} \quad (2.2.24)$$

$$= 2 \begin{bmatrix} \sum_{i=1}^b p_{i,1} & \cdots & 0 & \sum_{i=1}^b p_{i,1} x_i & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{i=1}^b p_{i,J} & 0 & \cdots & \sum_{i=1}^b p_{i,J} x_i \\ \sum_{i=1}^b p_{i,1} x_i & \cdots & 0 & \sum_{i=1}^b p_{i,1} x_i^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{i=1}^b p_{i,J} x_i & 0 & \cdots & \sum_{i=1}^b p_{i,J} x_i^2 \end{bmatrix}$$

$$Hess(S_2) = 2 \begin{bmatrix} \frac{\partial^2 S_2}{\partial x_1^2} & \cdots & \frac{\partial^2 S_2}{\partial x_1 \partial x_b} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 S_2}{\partial x_b \partial x_1} & \cdots & \frac{\partial^2 S_2}{\partial x_b^2} \end{bmatrix} = 2 \begin{bmatrix} \sum_{j=1}^J p_{1,j} \beta_j^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{j=1}^J p_{b,j} \beta_j^2 \end{bmatrix} \quad (2.2.25)$$

Estabelecamos a

Proposição 2.4 *As matrizes $Hess(S_j)$, $j = 1, 2$, são definidas positivas.*

Dem: Para a segunda destas matrizes a tese é evidente, pelo que nos basta considerar a primeira. Assim, a forma quadrática

$$\begin{aligned} \frac{1}{2} \mathbf{v}'_{2J} Hess(S_1) \mathbf{v}_{2J} &= \sum_{j=1}^J v_j^2 \sum_{i=1}^b p_{i,j} + 2 \sum_{j=1}^J v_j v_{J+j} \sum_{i=1}^b p_{i,j} x_i + \sum_{j=1}^J v_{J+j}^2 \sum_{i=1}^b p_{i,j} x_i^2 \\ &= \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (v_j^2 + 2v_j v_{J+j} x_i + v_{J+j}^2 x_i^2) \\ &= \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (v_j + v_{J+j} x_i)^2 > 0 \end{aligned} \quad (2.2.26)$$

é definida positiva para qualquer vector \mathbf{v}_{2J} não nulo, o que completa a demonstração. ■

Aliás, esta discussão em termos das matrizes hessianas, pode de certa forma, prolongar-se para a função objectivo global $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$. Para isso suponhamos que, após a realização de algumas iterações do algoritmo Zig-Zag, se obteve um bom ajustamento final, do qual resultaram os estimadores dos índices ambientais $\tilde{x}_i, i = 1, \dots, b$ e dos coeficientes de regressão $(\tilde{\alpha}_j, \tilde{\beta}_j), j = 1, \dots, J$. Consideremos ainda que a boa qualidade do ajustamento efectuado, nos permite aproximar os resíduos, que serão assim dados por,

$$r_{i,j} = p_{i,j}(y_{i,j} - \tilde{\alpha}_j - \tilde{\beta}_j \tilde{x}_i) \approx 0; i = 1, \dots, b; j = 1, \dots, J. \quad (2.2.27)$$

Devido ao resultado anterior, é possível escrever-mos as derivadas parciais $\frac{\partial^2 S}{\partial x_i \partial \beta_j}$ de forma bastante mais simplificada,

$$\frac{\partial^2 S}{\partial x_i \partial \beta_j} = -2p_{i,j}(y_{i,j} - \alpha_j - \beta_j x_i) + 2p_{i,j}\beta_j x_i \approx 2p_{i,j}\tilde{\beta}_j \tilde{x}_i, i = 1, \dots, b; j = 1, \dots, J. \quad (2.2.28)$$

Esta simplificação é muito útil, ver Pereira (2004), pois possibilita-nos aproximar a matriz hessiana de S pela matriz

$$\mathbf{A} = 2 \begin{bmatrix} \sum_{j=1}^J p_{1,j} \tilde{\beta}_j^2 & \cdots & 0 & p_{1,1} \tilde{\beta}_1 & \cdots & p_{1,J} \tilde{\beta}_J & p_{1,1} \tilde{\beta}_1 \tilde{x}_1 & \cdots & p_{1,J} \tilde{\beta}_J \tilde{x}_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{j=1}^J p_{b,j} \tilde{\beta}_j^2 & p_{b,1} \tilde{\beta}_1 & \cdots & p_{b,J} \tilde{\beta}_J & p_{b,1} \tilde{\beta}_1 \tilde{x}_b & \cdots & p_{b,J} \tilde{\beta}_J \tilde{x}_b \\ p_{1,1} \tilde{\beta}_1 & \cdots & p_{b,1} \tilde{\beta}_1 & \sum_{i=1}^b p_{i,1} & \cdots & 0 & \sum_{i=1}^b p_{i,1} \tilde{x}_i & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{1,J} \tilde{\beta}_J & \cdots & p_{b,J} \tilde{\beta}_J & 0 & \cdots & \sum_{i=1}^b p_{i,J} & 0 & \cdots & \sum_{i=1}^b p_{i,J} \tilde{x}_i \\ p_{1,1} \tilde{\beta}_1 \tilde{x}_1 & \cdots & p_{b,1} \tilde{\beta}_1 \tilde{x}_b & \sum_{i=1}^b p_{i,1} \tilde{x}_i & \cdots & 0 & \sum_{i=1}^b p_{i,1} \tilde{x}_i^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{1,J} \tilde{\beta}_J \tilde{x}_1 & \cdots & p_{b,J} \tilde{\beta}_J \tilde{x}_b & 0 & \cdots & \sum_{i=1}^b p_{i,J} \tilde{x}_i & 0 & \cdots & \sum_{i=1}^b p_{i,J} \tilde{x}_i^2 \end{bmatrix}.$$

Ora, com $\mathbf{u}'_{b+2J} = [\mathbf{a}'_b, \mathbf{b}'_J, \mathbf{c}'_J]$, virá

$$\begin{aligned}
\mathbf{u}'_{b+2J} \mathbf{A} \mathbf{u}_{b+2J} &= 2 \sum_{i=1}^b \left(\sum_{j=1}^J p_{i,j} \tilde{\beta}_j^2 \right) a_i^2 + 4 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \tilde{\beta}_j \tilde{x}_i a_i c_j + 4 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \tilde{\beta}_j a_i b_j \\
&+ 2 \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \right) b_j^2 + 2 \left[2 \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \tilde{x}_i \right) b_j c_j + \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \tilde{x}_i^2 \right) c_j^2 \right] \\
&= 2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \left(\tilde{\beta}_j^2 a_i^2 + 2 \tilde{\beta}_j a_i (b_j + c_j \tilde{x}_i) + b_j^2 + 2 b_j c_j \tilde{x}_i + c_j^2 \tilde{x}_i^2 \right) \\
&= 2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \left[\tilde{\beta}_j a_i + (b_j + c_j \tilde{x}_i) \right]^2 > 0
\end{aligned} \tag{2.2.29}$$

o que nos permite dizer que, numa vizinhança do óptimo, podemos considerar que $Hess(S)$ é uma matriz definida positiva, o que nos leva a antecipar bom comportamento para o método.

2.2.3 Redes Fraccionalmente completas

Suponhamos que temos uma rede de ensaios que se decompõe naturalmente em sub-redes conectadas. Em particular, interessa-nos o caso em que cada sub-rede corresponde a um ano. Assim, estamos a considerar um plano composto por várias sub-redes [vários anos], havendo para cada sub-rede um conjunto de locais e um conjunto de cultivares utilizados. Dizemos que numa sub-rede se tem o caso completo, quando o conjunto dos cultivares utilizados nessa sub-rede é o mesmo para todos os locais da sub-rede. Consideramos ainda que, se em cada uma dessas sub-redes se tiver o caso completo, temos uma rede fraccionalmente completa. Veja-se que, como o conjunto de cultivares pode variar de sub-rede para sub-rede, uma rede pode ser fraccionalmente completa sem ser completa.

Tendo-se ordenado as sub-redes numa rede global de 1 a L , interessa-nos estabelecer a

Proposição 2.5 *Se, tendo-se ordenado as sub-redes de 1 a L , sub-redes contíguas tiverem cultivares em comum, a rede global será conectada.*

Dem: Dados os blocos $b_i[b'_i]$ pertencentes às sub-redes $r_i[r'_i]$, para construir uma cadeia indo do primeiro para o segundo podemos construir uma cadeia que passe sucessivamente em todas as sub-redes intermédias. Desde que as sub-redes sejam conectadas, todos os pares de sub-redes intermédias têm um cultivar em comum, havendo sempre um par de blocos, cada bloco pertencente a uma sub-rede, contendo um cultivar em comum, o que estabelece a tese. ■

A proposição seguinte tem interesse, pois como veremos mais adiante neste trabalho, no caso de redes fraccionalmente completas conectadas, existe conexão na rede global e pode proceder-se a um ajustamento conjunto a par do ajustamento sub-rede a sub-rede. Designaremos o ajustamento conjunto por Opção *I* e os ajustamentos parciais por Opção *II*. Na aplicação que vamos apresentar, como iremos ver, temos uma sub-rede para cada ano, sendo a rede global fraccionalmente completa.

Sejam S_1, \dots, S_L as funções objectivo das sub-redes numa rede fraccionalmente completa e S a respectiva função objectivo da rede global. Representemos por $S^*[S_1^*, \dots, S_L^*]$ os mínimos da função objectivo global [das funções objectivo parciais]. Estabeleçamos a

Proposição 2.6 *Tem-se $S^* > \sum_{l=1}^L S_l^*$.*

Dem: Como cada bloco pertence a uma e só uma das sub-redes ter-se-á

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{l=1}^L S_l(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \quad (2.2.30)$$

com

$S_l(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{i \in B_l} \sum_{j=1}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2$ sendo $B_l, l = 1, \dots, L$ o conjunto dos índices dos blocos da l -ésima sub-rede. Se um cultivar não pertence à sub-rede l tem-se $p_{i,j} = 0$ para $i \in B_l$. Ora, ao obtermos os mínimos $S_l^*, l = 1, \dots, L$, apenas temos de atender ao que se passa para os blocos das sub-redes correspondentes. Assim, se um cultivar pertencer a duas ou mais sub-redes, poderão os estimadores dos respectivos coeficientes variar de sub-rede para sub-rede. Pelo contrário, ao minimizarmos S , tem de se introduzir a restrição desses coeficientes serem iguais. Desta forma, verifica-se que o domínio em que se minimiza S é mais restrito do que o que se tem quando se realizam minimizações separadas. O resto da demonstração é directa. ■

2.3 Inferência

2.3.1 Algumas considerações preliminares

Apresentamos nesta secção sobre inferência, algumas considerações e pressupostos de base, bem como os resultados teóricos necessários à interpretação dum plano de melhoramento com determinadas características. Veremos a aplicação prática dos referidos resultados, quando no terceiro capítulo, procedermos à análise pormenorizada de um plano de melhoramento concreto.

Assim começamos por admitir, como pressuposto de base, que o vector dos rendimentos para os cultivares, \mathbf{y}_n , tem distribuição normal, sendo homocedástico, com vector médio $\mathbf{X}(\boldsymbol{\theta}_b)\boldsymbol{\gamma}_{2J}$. Observe-se que pressupostos deste tipo, relativos a variáveis bio-métricas, foram introduzidos inicialmente por Fisher (1918). Neste trabalho, Fisher aponta para a existência dum grande número de pares de alelos intervenientes, o que neste contexto se pode interpretar como sendo uma versão física do teorema do limite central.

A inferência que vamos fazer, surge na sequência dos resultados obtidos pelo ajustamento do modelo definido na subsecção 2.2.2. Recorde-se que, o ajustamento pode ser efectuado globalmente, pela Opção *I*, ou parcialmente pela Opção *II*. Mais precisamente faremos inferência sobre:

a) coeficientes individuais;

b) funções lineares da forma $U_{j,l}(u, v) = \alpha_{j,l}u + \beta_{j,l}v, j = 1, \dots, J; l = 1, \dots, h$.

Apresentaremos um primeiro parágrafo sobre questões relativas às funções lineares. Neste caracterizaremos as funções objectivo lineares, estabelecendo a sua distribuição estatística, que como veremos, se distingue para cada uma das Opções. Veremos ainda a importância de seleccionarmos funções objectivo invariantes à mudança de escala. Seguem-se dois parágrafos em que se testam hipóteses relativas às mesmas, primeiro quando se considera um factor, cultivar ou ano, e depois quando se consideram ambos e a respectiva interacção.

Na sua dissertação de doutoramento, Pereira (2004) mostrou que, se se admitir a normalidade e homocedasticidade dos rendimentos, os estimadores obtidos utilizando índices ambientais L_2 , são estimadores de máxima verosimilhança. Apresentou ainda, ver Pereira (2004), uma bateria de testes de quociente de verosimilhanças, para hipóteses sobre os coeficientes de regressão. Dada a importância teórica desta matéria, procuraremos na subsecção 2.3.5 completar essa bateria de testes. Por último mostraremos, na subsecção 2.3.6, como obter regiões de confiança para os vectores dos coeficientes de regressão e dos índices ambientais estimados.

2.3.2 Funções Objectivo Lineares

As análises comparativas que faremos, incidirão nos valores de funções lineares do tipo,

$$U_j(u, v) = \alpha_j u + \beta_j v, j = 1, \dots, J \quad (2.3.1)$$

onde o parâmetro u tomará usualmente o valor unitário, e o parâmetro v tomará valores no intervalo de variação admitido para os índices ambientais. Após o ajustamento do modelo com o algoritmo Zig-Zag, obter-se-á para cada cultivar j a função estimada,

$$\tilde{U}_j(u, v) = \tilde{\alpha}_j u + \tilde{\beta}_j v, j = 1, \dots, J. \quad (2.3.2)$$

Trataremos estas funções como variáveis aleatórias com distribuição aproximadamente Normal. No parágrafo seguinte, caracterizaremos a sua distribuição particularizando os casos em que o ajustamento é realizado no âmbito da Opção *I* ou da Opção *II*.

Opção *I*/Opção *II*

Começemos por abordar o caso em que as funções lineares são estimadas, utilizando os resultados da Opção *I* e posteriormente trataremos o caso em que as mesmas são estimadas com os resultados da Opção *II*.

Suponhamos a existência de uma rede global, conectada fraccionalmente completa, correspondendo as sub-redes a diferentes anos. Considere-se ainda, que após a realização de um ajustamento global, se obtiveram a partir da última iteração do algoritmo Zig-zag, os seguintes estimadores:

1. o índice ambiental ajustado mínimo, \tilde{x}^- ;
2. o índice ambiental ajustado máximo, \tilde{x}^+ ;
3. por cultivar o índice ambiental ajustado médio, $\tilde{x}_j, j = 1, \dots, J$;
4. a soma dos quadrados dos resíduos para a média dos índices ambientais ajustados para o cultivar j , $\tilde{s}_j, j = 1, \dots, J$;
5. a soma das somas dos quadrados dos resíduos para as regressões ajustadas na última iteração do algoritmo Zig-Zag, \tilde{S} ;
6. os coeficientes ajustados,

$$\begin{bmatrix} \tilde{\alpha}_j \\ \tilde{\beta}_j \end{bmatrix} \approx N \left(\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{1}{b_j} + \frac{x_{.j}^2}{s_j} & -\frac{x_{.j}}{s_j} \\ -\frac{x_{.j}}{s_j} & \frac{1}{s_j} \end{bmatrix} \right), j = 1, \dots, J \quad (2.3.3)$$

independentes de \tilde{S} , que se distribui como o produto de σ^2 por um qui-quadrado com $g = n - b - 2J$ graus de liberdade, pondo-se $\tilde{S} \sim \sigma^2 \chi_g^2$. Para aligeirar a escrita, ao longo do trabalho escreveremos S em vez de \tilde{S} .

Note-se que, para o cultivar j , o índice ambiental ajustado médio \tilde{x}_j , é obtido calculando a média dos índices ambientais estimados para os b_j blocos em que o cultivar j está presente e \tilde{s}_j será a soma dos quadrados dos desvios desses índices ambientais para essa média.

A nossa abordagem integra-se na dos delineamentos regressionais múltiplos, ver Mexia (1987). Nestes delineamentos, ajusta-se uma regressão linear, nas mesmas variáveis, para

cada um dos tratamentos dum delineamento base. O erro é estimado a partir de S . Em geral, testam-se hipóteses sobre combinações lineares dos coeficientes das regressões, podendo ter-se por exemplo os casos particulares, $U_j(1, 0) = \alpha_j$ ou $U_j(0, 1) = \beta_j, j = 1, \dots, J$.

Utilizando os estimadores obtidos pela Opção *I*, temos as funções estimadas $\tilde{U}_j(u, v) \approx N[\theta_j(u, v), \sigma^2 K_j(u, v)], j = 1, \dots, J$, com

$$\theta_j(u, v) = \alpha_j u + \beta_j v \quad (2.3.4)$$

e

$$K_j(u, v) = \frac{u^2}{b_j} + \frac{(ux_{.j} - v)^2}{s_j}, j = 1, \dots, J. \quad (2.3.5)$$

Vejamos agora o que se passa, quando utilizamos os estimadores obtidos pela Opção *II*. Seja C_l o conjunto dos índices dos cultivares presentes no ano l . Suponhamos que os cultivares com índices $j \in C_l$, estão presentes nos anos com índices $l = 1, \dots, h$ e que a realização de um ajustamento ano a ano deu origem aos estimadores para o índice ambiental mínimo \tilde{x}_l^- , máximo \tilde{x}_l^+ , médio \tilde{x}_l e para as somas dos quadrados dos resíduos para as médias $\tilde{s}_l, l = 1, \dots, h$. Estando o cultivar j presente no ano l , sejam $(\tilde{\alpha}_{j,l}, \tilde{\beta}_{j,l})$ os coeficientes ajustados para esse ano. Teremos assim as funções estimadas para o j -ésimo cultivar e l -ésimo ano,

$$\tilde{U}_{j,l}(u, v) = \tilde{\alpha}_{j,l}u + \tilde{\beta}_{j,l}v \approx N[\theta_{j,l}(u, v), \sigma^2 K_{j,l}(u, v)], j \in C_l, l = 1, \dots, h. \quad (2.3.6)$$

Para um dado ano, como referimos, estamos a considerar o caso completo, em que todos os m cultivares estão presentes em todos os b blocos, tendo-se assim, $x_{.1} = \dots = x_{.m} = x$ e $s_1 = \dots = s_m = s$, vindo

$$K_j(u, v) = K(u, v) = \frac{u^2}{b} + \frac{(ux_{.j} - v)^2}{s}, j \in C_l, \quad (2.3.7)$$

o que simplifica os cálculos.

O problema que se coloca ao realizar um ajustamento para cada ano, é que o intervalo de variação para os índices ambientais virá muito provavelmente distinto em cada ano. Para algumas das análises comparativas que vamos fazer, pode ser conveniente trabalharmos sempre com o mesmo intervalo de variação para os índices ambientais. Para tal podemos, ano a ano, estandardizar o intervalo de variação dos índices ambientais, por forma a que, na vez dos intervalos $[\tilde{x}_l^-, \tilde{x}_l^+], l = 1, \dots, h$, se tenha para todos os anos um intervalo de extremos comuns dados por \tilde{x}^- e \tilde{x}^+ . Com $\tilde{x} \in [\tilde{x}_l^-, \tilde{x}_l^+]$, a estandardização levará à transformação

$$\tilde{x}' = c_l \tilde{x} + d_l \quad (2.3.8)$$

com

$$\begin{cases} c_l &= \frac{\tilde{x}^+ - \tilde{x}^-}{\tilde{x}_l^+ - \tilde{x}_l^-} \\ d_l &= \tilde{x}^- - c_l \tilde{x}_l^- \end{cases} \quad (2.3.9)$$

Note-se que por (2.2.18), os coeficientes ajustados para o j -ésimo cultivar serão agora,

$$\begin{cases} \tilde{\beta}'_{j,l} &= \frac{\tilde{\beta}_{j,l}}{c_l} \\ \tilde{\alpha}'_{j,l} &= \tilde{\alpha}_{j,l} - \frac{d_l}{c_l} \tilde{\beta}_{j,l} \end{cases} \quad (2.3.10)$$

É conveniente utilizarmos para a inferência, funções lineares com valores invariantes para estas transformações. Nesse caso, garante-se que as transformações devidas ao processo de standardização não influenciam em nada os valores tomados por estas funções. Para clarificar o que acabámos de ver, estabeleçamos a

Proposição 2.7 *A função objectivo $\tilde{U}_{j,l}(1, x) = \tilde{\alpha}_{j,l} + \tilde{\beta}_{j,l}x$ toma valores invariantes à mudança de escala nos pontos $x = r\tilde{x}^+ + (1 - r)\tilde{x}^-$, com $0 < r < 1$.*

Dem: Com $0 < r < 1$ e $x = r\tilde{x}^+ + (1 - r)\tilde{x}^-$, ter-se-á antes da standardização

$$\tilde{U}_{j,l}(1, x) = \tilde{\alpha}_{j,l} + \tilde{\beta}_{j,l}x. \quad (2.3.11)$$

A standardização como se viu por (2.2.18) e (2.2.19) leva às seguintes transformações:

$$\begin{cases} \tilde{\beta}'_{j,l} &= \frac{\tilde{\beta}_{j,l}}{c_l} \\ \tilde{\alpha}'_{j,l} &= \tilde{\alpha}_{j,l} - \frac{d_l}{c_l} \tilde{\beta}_{j,l} \\ x' &= c_l x + d_l = c_l[r\tilde{x}^+ + (1 - r)\tilde{x}^-] + d_l \end{cases} \quad (2.3.12)$$

ter-se-á então, após a standardização,

$$\tilde{U}'_{j,l}(1, x') = \tilde{\alpha}'_{j,l} + \tilde{\beta}'_{j,l}x' = \tilde{\alpha}_{j,l} - \frac{d_l}{c_l} \tilde{\beta}_{j,l} + \frac{1}{c_l} \tilde{\beta}_{j,l}x', \quad (2.3.13)$$

substituindo x' dado por (2.3.12), na expressão anterior obtém-se,

$$\tilde{U}'_{j,l}(1, x') = \tilde{\alpha}_{j,l} + \tilde{\beta}_{j,l}r\tilde{x}^+ + \tilde{\beta}_{j,l}\tilde{x}^- - \tilde{\beta}_{j,l}r\tilde{x}^- = \tilde{\alpha}_{j,l} + \tilde{\beta}_{j,l}x \quad (2.3.14)$$

o que estabelece a tese. ■

Salientamos ainda que, as comparações que adiante faremos, situam-se no caso equilibrado. Ora, a análise de variância e técnicas relacionadas são, ver Ito (1980), robustas nesse caso, o que nos permite aplicar as mesmas, embora os valores das funções objectivo sejam apenas aproximadamente normais.

2.3.3 Hipóteses sobre um factor

Vamos começar por apresentar um teste de hipóteses, que utilizaremos adiante quando abordarmos a questão da selecção dos cultivares. Este teste, destina-se a comparar para um determinado valor v_o do índice ambiental, uma ou mais das funções lineares $U_j(u, v)$, com uma delas, $U_{j'}(u, v)$, tomada como padrão. Com $u = 1$ e $v = v_o$, teremos para o par (j, j') as hipóteses testada e alternativa

$$\begin{cases} H_{0,j,j'}(1, v_o) : U_j(1, v_o) = U_{j'}(1, v_o) \\ H_{1,j,j'}(1, v_o) : U_j(1, v_o) > U_{j'}(1, v_o) \end{cases} \quad (2.3.15)$$

com

$$U_j(1, v_o) = \alpha_j + \beta_j v_o, \quad j, j' = 1, \dots, J \text{ e } j \neq j'.$$

A estatística de teste a utilizar é,

$$t_{j,j'}(1, v_o) = \frac{\tilde{U}_j(1, v_o) - \tilde{U}_{j'}(1, v_o)}{\sqrt{\frac{S}{g}[K_j(1, v_o) + K_{j'}(1, v_o)]}} \quad (2.3.16)$$

com, quando H_o se verifica, distribuição t central com g graus de liberdade.

Podemos ainda utilizar os métodos de comparação múltipla de Scheffé e de Bonferroni para comparar as $U_j(1, v_o)$, $j = 1, \dots, J$.

Assim, ver Scheffé (1959, págs. 66 a 70), $\tilde{U}_{j'}(1, v_o) = \tilde{\alpha}_{j'} + \tilde{\beta}_{j'} v_o$ será, ao nível α , conjunta e significativamente superior aos $\tilde{U}_j(1, v_o) = \tilde{\alpha}_j + \tilde{\beta}_j v_o$ tais que $j < j'$ e que

$$\frac{\tilde{U}_{j'}(1, v_o) - \tilde{U}_j(1, v_o)}{\sqrt{[K_j(1, v_o) + K_{j'}(1, v_o)](j' - 1) \cdot F_{1-\alpha, j'-1, g} \frac{S}{g}}} > 1, \quad (2.3.17)$$

e aos tais que $j > j'$ e que

$$\frac{\tilde{U}_{j'}(1, v_o) - \tilde{U}_j(1, v_o)}{\sqrt{[K_j(1, v_o) + K_{j'}(1, v_o)](J - j') \cdot F_{1-\alpha, J-j', g} \frac{S}{g}}} > 1, \quad (2.3.18)$$

sendo $F_{1-\alpha, v_1, v_2}$ o valor crítico ao nível α da distribuição F de Snedcor central com v_1 e v_2 graus de liberdade.

Quando o nosso interesse é inferir sobre um número fixo de contrastes, podemos usar a desigualdade de Bonferroni para melhorar os resultados obtidos com o método de comparação múltipla de Scheffé. Quando nos interessa apenas comparar pares de valores, podemos ver Seber (2003, pág. 120:125), substituir nas expressões anteriores $\sqrt{(j' - 1) \cdot F_{1-\alpha, j'-1, g}}$ e $\sqrt{(J - j') \cdot F_{1-\alpha, J-j', g}}$ pelos quantis da distribuição t de Student com g graus de liberdade,

para as probabilidades $1 - \frac{\alpha}{2 \cdot (j'-1)}$ e $1 - \frac{\alpha}{2 \cdot (J-j')}$ respectivamente. Estes testes e aplicações de métodos de comparação múltipla, ser-nos-ão úteis quando, ao estudarmos o contorno superior duma ACR, compararmos cultivares.

Do ponto de vista da condução de um plano de melhoramento, valoriza-se geralmente a obtenção de cultivares com estabilidade de resposta às variações dos índices ambientais. Neste contexto, passamos agora à apresentação de outro teste, que permite verificar a existência de estabilidade no comportamento dos cultivares, ao longo dos anos em que os mesmos são utilizados. Tal estabilidade, traduzir-se-á para um cultivar, pela inexistência de diferenças significativas nos coeficientes angulares das regressões ajustadas, em diferentes anos, para esse cultivar. Diremos que, um cultivar que satisfaz esta condição, é β -estável. Centra-mo-nos nos coeficientes angulares, dada a grande variabilidade dos estimadores das ordenadas na origem. Teremos agora modelos em que varia o factor ano.

Seja $D_j = (u_1, \dots, u_{m_j})$ o conjunto dos índices dos anos em que o j -ésimo cultivar foi utilizado. Para cada cultivar $j, j = 1, \dots, J$, vamos comparar os valores estimados para as funções,

$$\tilde{U}_{j,l}(0, 1) = \tilde{\beta}_{j,l}, l = u_1, \dots, u_{m_j} \quad (2.3.19)$$

o que podemos fazer testando as hipóteses,

$$H_{o,j} : \beta_{j,u_1} = \dots = \beta_{j,u_{m_j}}, j = 1, \dots, J. \quad (2.3.20)$$

Antes de prosseguirmos, vamos estabelecer alguns resultados úteis à obtenção da estatística de teste para a hipótese acima enunciada. Começamos pelo

Lema 2.3 *Sendo \mathbf{A} a matriz que se obtém, retirando a primeira linha igual a $\frac{1}{\sqrt{m}}\mathbf{1}_m$ a uma matriz ortogonal, e \mathbf{K} uma matriz definida positiva, as matrizes \mathbf{AKA}' e $(\mathbf{AKA}')^{-1}$ são definidas positivas.*

Dem: Como \mathbf{K} é definida positiva, pode escrever-se $\mathbf{K} = \mathbf{P}' \mathbf{diag}(r_1, \dots, r_m) \mathbf{P}$ com \mathbf{P} matriz ortogonal e $r_i > 0, i = 1, \dots, m$, os valores próprios positivos de \mathbf{K} . Assim, estará definida a matriz $\mathbf{K}^{\frac{1}{2}} = \mathbf{P}' \mathbf{diag}(r_1^{\frac{1}{2}}, \dots, r_m^{\frac{1}{2}}) \mathbf{P}$, também definida positiva e tal que $\mathbf{K} = \mathbf{K}^{\frac{1}{2}} \mathbf{K}^{\frac{1}{2}}$, vindo $\mathbf{AKA}' = \mathbf{AK}^{\frac{1}{2}} \mathbf{K}^{\frac{1}{2}} \mathbf{A}' = (\mathbf{AK}^{\frac{1}{2}})(\mathbf{AK}^{\frac{1}{2}})'$. Como $\mathbf{K}^{\frac{1}{2}}$ é regular ter-se-á, ver Silvey (1975, pág. 176), $Car(\mathbf{AKA}') = Car(\mathbf{AK}^{\frac{1}{2}}) = Car(\mathbf{A}) = m - 1$, pelo que \mathbf{AKA}' será regular e como $\mathbf{v}'(\mathbf{AKA}')\mathbf{v} = \|\mathbf{AK}^{\frac{1}{2}}\mathbf{v}\|^2 > 0$, qualquer que seja $\mathbf{v} \neq \mathbf{0}$, \mathbf{AKA}' será definida positiva. Se uma matriz é definida positiva, ver Judge (1988, pág. 961), a sua inversa também o é, completando-se assim a demonstração do lema. ■

Proposição 2.8 Dado $\mathbf{Z}_m \sim N(\boldsymbol{\eta}_m, \sigma^2 \mathbf{K})$ com \mathbf{K} definida positiva, independente de $S \sim (\sigma^2 \chi_g^2)$ e \mathbf{A} a matriz obtida retirando a primeira linha igual a $\frac{1}{\sqrt{m}} \mathbf{1}_m$ a uma matriz ortogonal, a estatística

$$F = \frac{g}{m-1} \cdot \frac{(\mathbf{AZ}_m)'(\mathbf{AKA}')^{-1}\mathbf{AZ}_m}{S} \quad (2.3.21)$$

tem distribuição F de Snedcor com $m-1$ e g graus de liberdade e parâmetro de não centralidade $\delta = \frac{1}{\sigma^2}(\mathbf{A}\boldsymbol{\eta}_m)'(\mathbf{AKA}')^{-1}(\mathbf{A}\boldsymbol{\eta}_m)$ que se anula se e só se as componentes de $\boldsymbol{\eta}_m$ são idênticas.

Dem: Tem-se, veja-se Seber (1980, pág. 5), $\mathbf{AZ}_m \sim N(\mathbf{A}\boldsymbol{\eta}_m, \sigma^2 \mathbf{AKA}')$, vindo, ver Mexia (1995, pág. 52), $(\mathbf{AZ}_m)'(\mathbf{AKA}')^{-1}\mathbf{AZ}_m \sim \sigma^2 \chi_{m-1, \delta}^2$ independente de $S \sim \sigma^2 \chi_g^2$, donde resulta a primeira parte da tese. Atendendo ao Lema 2.3, \mathbf{AKA}' é definida positiva, pelo que podemos escrever, $\mathbf{AKA}' = \dot{\mathbf{P}}' \mathbf{diag}(s_1, \dots, s_{m-1}) \dot{\mathbf{P}}$, com $\dot{\mathbf{P}}$ matriz ortogonal sendo os $s_i > 0, i = 1, \dots, m-1$, os valores próprios de \mathbf{AKA}' . Ter-se-á então, $(\mathbf{AKA}')^{-1} = \dot{\mathbf{P}}' \mathbf{diag}(s_1^{-1}, \dots, s_{m-1}^{-1}) \dot{\mathbf{P}}$, vindo $\delta = 0$ se e só se $\dot{\mathbf{P}}\mathbf{A}\boldsymbol{\eta}_m = \mathbf{0}_{m-1}$, o que, dado $\dot{\mathbf{P}}$ ser regular, equivale a ter-se $\mathbf{A}\boldsymbol{\eta}_m = \mathbf{0}_{m-1}$. Ora, \mathbf{A} é constituída por $m-1$ linhas duma matriz ortogonal de ordem m , tendo-se pois $\text{Car}(\mathbf{A}) = m-1$ e sendo $N(\mathbf{A})$ o espaço de nulidade de \mathbf{A} , $\dim(N(\mathbf{A})) = m - \text{Car}(\mathbf{A}) = 1$. Basta agora observar que por construção, $\frac{1}{\sqrt{m}} \mathbf{1}_m$ é ortogonal aos vectores linha de \mathbf{A} , para se ver que os vectores de $N(\mathbf{A})$ são do tipo $\frac{c}{\sqrt{m}} \mathbf{1}_m$. Assim teremos, $\delta = 0$ se e só se $\boldsymbol{\eta}_m = \frac{c}{\sqrt{m}} \mathbf{1}_m$, o que estabelece a segunda parte da tese. ■

Neste caso, estamos a trabalhar com os coeficientes ajustados em cada ano, portanto estamos a utilizar os resultados da Opção II. Assim, sendo s_{u_l} a soma dos quadrados dos desvios para a média dos índices ambientais ajustados no ano com índice $u_l, l = 1, \dots, m_j$, para o j -ésimo cultivar tomamos,

$$\mathbf{K} = \mathbf{diag}\left(\frac{1}{s_{u_1}}, \dots, \frac{1}{s_{u_{m_j}}}\right) \quad (2.3.22)$$

bem como

$$\mathbf{Z}'_m = [\tilde{\beta}_{j,u_1}, \dots, \tilde{\beta}_{j,u_{m_j}}]. \quad (2.3.23)$$

2.3.4 Hipóteses sobre dois factores com interacção

Vamos considerar uma situação semelhante à análise de variância com dois factores cruzados com interacção. No nosso caso, os níveis dos factores são cultivares e anos. Aos pares (cultivar, ano) correspondem os pares (j, l) de índices com $j = 1, \dots, m$ e $l = 1, \dots, h$. Teremos então para as funções estimadas,

$$\tilde{U}_{j,l}(1, x) = \tilde{\alpha}_{j,l} + \tilde{\beta}_{j,l}x, j = 1, \dots, m; l = 1, \dots, h \quad (2.3.24)$$

os valores médios

$$U_{j,l}(1, x) = \alpha_{j,l} + \beta_{j,l}x, j = 1, \dots, m; l = 1, \dots, h. \quad (2.3.25)$$

Tomando-se

$$\begin{cases} U_{j\cdot} = \frac{1}{h} \sum_{l=1}^h U_{j,l}, j = 1, \dots, m \\ U_{\cdot,l} = \frac{1}{m} \sum_{j=1}^m U_{j,l}, l = 1, \dots, h \\ U_{\cdot\cdot} = \frac{1}{m \cdot h} \sum_{j=1}^m \sum_{l=1}^h U_{j,l} \end{cases} \quad (2.3.26)$$

teremos os efeitos dos níveis dos dois factores e as interacções entre pares de níveis dadas por,

$$\begin{cases} \gamma_j = U_{j\cdot} - U_{\cdot\cdot}, & j = 1, \dots, m \\ \eta_l = U_{\cdot,l} - U_{\cdot\cdot}, & l = 1, \dots, h \\ \lambda_{j,l} = U_{j,l} - U_{j\cdot} - U_{\cdot,l} + U_{\cdot\cdot}, & j = 1, \dots, m; l = 1, \dots, h \end{cases} \quad (2.3.27)$$

Note-se que, embora se disponha dum único valor por cada par de níveis, não somos forçados a admitir a aditividade, pois os valores estimados das funções $\tilde{U}_{j,l}(1, x)$ são independentes da soma S das somas dos quadrados dos resíduos para as regressões ajustadas. Por ser mais conveniente, combinamos os índices j e l num índice único,

$$i = l + (j - 1)h \quad (2.3.28)$$

criando-se um novo vector \tilde{U}_w , cujas componentes com índices $i = 1, \dots, w = mh$ virão assim agrupadas de acordo com os cultivares. O vector médio U_w de \tilde{U}_w variará na totalidade do espaço \mathfrak{R}_w .

À semelhança da análise de variância com dois factores, interessa-nos testar as hipóteses de ausência de efeitos

$$\begin{cases} H_{o,\gamma} : \gamma_j = 0; j = 1, \dots, m \\ H_{o,\eta} : \eta_l = 0; l = 1, \dots, h, \end{cases} \quad (2.3.29)$$

e de ausência de interacções

$$H_{o,\lambda} : \lambda_{j,l} = 0; j = 1, \dots, m; l = 1, \dots, h. \quad (2.3.30)$$

Seguindo Mexia (1989), vamos reformular estas hipóteses em termos dos espaços de nulidade das matrizes \mathbf{A}_γ , \mathbf{A}_η e \mathbf{A}_λ , escrevendo-se,

$$\begin{cases} H_{o,\gamma} : \mathbf{A}_\gamma \mathbf{U}_w = \mathbf{0}_{m-1} \\ H_{o,\eta} : \mathbf{A}_\eta \mathbf{U}_w = \mathbf{0}_{h-1} \\ H_{o,\lambda} : \mathbf{A}_\lambda \mathbf{U}_w = \mathbf{0}_{(m-1)(h-1)} \end{cases} \quad (2.3.31)$$

O produto de Kronecker da matriz $\mathbf{C} = [c_{i,j}]$ do tipo $(r \times s)$ pela matriz \mathbf{B} , dado por

$$\mathbf{C} \otimes \mathbf{B} = \begin{bmatrix} c_{1,1}\mathbf{B} & \cdots & c_{1,s}\mathbf{B} \\ \vdots & \cdots & \vdots \\ c_{r,1}\mathbf{B} & \cdots & c_{r,s}\mathbf{B} \end{bmatrix} \quad (2.3.32)$$

é útil na obtenção das matrizes \mathbf{A}_γ , \mathbf{A}_η e \mathbf{A}_λ .

Dado haver cruzamento dos dois factores tem-se, ver Fonseca et al (2003),

$$\begin{cases} \mathbf{A}_\gamma = \mathbf{L}_m \otimes \frac{1}{\sqrt{h}}\mathbf{1}'_h \\ \mathbf{A}_\eta = \frac{1}{\sqrt{m}}\mathbf{1}'_m \otimes \mathbf{L}_h \\ \mathbf{A}_\lambda = \mathbf{L}_m \otimes \mathbf{L}_h \end{cases} \quad (2.3.33)$$

com \mathbf{L}_m e \mathbf{L}_h obtidas retirando as primeiras linhas iguais a $\frac{1}{\sqrt{m}}\mathbf{1}'_m$ e $\frac{1}{\sqrt{h}}\mathbf{1}'_h$ a matrizes ortogonais com ordens m e h , respectivamente.

Por outro lado,

$$Var(\tilde{U}_{j,l}) = \sigma^2 K_l, j = 1, \dots, m; l = 1, \dots, h \quad (2.3.34)$$

com

$$K_l = \frac{1}{b_l} + \frac{(x - x_{..l})^2}{s_l}, l = 1, \dots, h. \quad (2.3.35)$$

Veja-se que através do ajustamento se obtém, \tilde{s}_l , a soma dos quadrados dos desvios dos índices ambientais do l -ésimo ano relativamente à respectiva média $\tilde{x}_{..l}$.

Assim, atendendo ao agrupamento das observações de acordo com os cultivares, o vector $\tilde{\mathbf{U}}_w$ terá matriz de covariância

$$Cov(\tilde{\mathbf{U}}_w) = \sigma^2 \mathbf{D} \quad (2.3.36)$$

com $\mathbf{D} = \mathbf{I}_m \otimes \mathbf{D}'$, sendo $\mathbf{D}' = \mathbf{diag}(K_1, \dots, K_h)$ a matriz diagonal com elementos principais K_1, \dots, K_h .

Seguindo os nossos pressupostos, teremos $\tilde{\mathbf{U}}_w \sim N(\mathbf{U}_w, \sigma^2 \mathbf{D})$, bem como,

$$\begin{cases} \mathbf{A}_\gamma \tilde{\mathbf{U}}_w \sim N(\mathbf{A}_\gamma \mathbf{U}_w, \sigma^2 \mathbf{A}_\gamma \mathbf{D} \mathbf{A}'_\gamma) \\ \mathbf{A}_\eta \tilde{\mathbf{U}}_w \sim N(\mathbf{A}_\eta \mathbf{U}_w, \sigma^2 \mathbf{A}_\eta \mathbf{D} \mathbf{A}'_\eta) \\ \mathbf{A}_\lambda \tilde{\mathbf{U}}_w \sim N(\mathbf{A}_\lambda \mathbf{U}_w, \sigma^2 \mathbf{A}_\lambda \mathbf{D} \mathbf{A}'_\lambda) \end{cases} \quad (2.3.37)$$

Como as matrizes de covariância dadas em (2.3.37) são simétricas e definidas positivas, teremos, ver por exemplo Mexia (1990),

$$\begin{cases} (\mathbf{A}_\gamma \tilde{\mathbf{U}}_w)' (\mathbf{A}_\gamma \mathbf{D} \mathbf{A}'_\gamma)^{-1} (\mathbf{A}_\gamma \tilde{\mathbf{U}}_w) \sim \sigma^2 \chi^2_{m-1, \delta_\gamma} \\ (\mathbf{A}_\eta \tilde{\mathbf{U}}_w)' (\mathbf{A}_\eta \mathbf{D} \mathbf{A}'_\eta)^{-1} (\mathbf{A}_\eta \tilde{\mathbf{U}}_w) \sim \sigma^2 \chi^2_{h-1, \delta_\eta} \\ (\mathbf{A}_\lambda \tilde{\mathbf{U}}_w)' (\mathbf{A}_\lambda \mathbf{D} \mathbf{A}'_\lambda)^{-1} (\mathbf{A}_\lambda \tilde{\mathbf{U}}_w) \sim \sigma^2 \chi^2_{(m-1)(h-1), \delta_\lambda} \end{cases} \quad (2.3.38)$$

com os parâmetros de não centralidade,

$$\begin{cases} \delta_\gamma &= \frac{1}{\sigma^2}(\mathbf{A}_\gamma \mathbf{U}_w)'(\mathbf{A}_\gamma \mathbf{D} \mathbf{A}_\gamma')^{-1}(\mathbf{A}_\gamma \mathbf{U}_w) \\ \delta_\eta &= \frac{1}{\sigma^2}(\mathbf{A}_\eta \mathbf{U}_w)'(\mathbf{A}_\eta \mathbf{D} \mathbf{A}_\eta')^{-1}(\mathbf{A}_\eta \mathbf{U}_w) \\ \delta_\lambda &= \frac{1}{\sigma^2}(\mathbf{A}_\lambda \mathbf{U}_w)'(\mathbf{A}_\lambda \mathbf{D} \mathbf{A}_\lambda')^{-1}(\mathbf{A}_\lambda \mathbf{U}_w) \end{cases} \quad (2.3.39)$$

que se anulam, quando e só quando as correspondentes hipóteses se verificam.

Dado \tilde{U}_w ser independente de $S \sim \sigma^2 \chi_g^2$, têm-se as estatísticas F de teste

$$\begin{cases} F_\gamma &= \frac{g}{m-1} \frac{(\mathbf{A}_\gamma \tilde{U}_w)'(\mathbf{A}_\gamma \mathbf{D} \mathbf{A}_\gamma')^{-1}(\mathbf{A}_\gamma \tilde{U}_w)}{S} \sim F(u|m-1, g, \delta_\gamma) \\ F_\eta &= \frac{g}{h-1} \frac{(\mathbf{A}_\eta \tilde{U}_w)'(\mathbf{A}_\eta \mathbf{D} \mathbf{A}_\eta')^{-1}(\mathbf{A}_\eta \tilde{U}_w)}{S} \sim F(u|h-1, g, \delta_\eta) \\ F_\lambda &= \frac{g}{(m-1)(h-1)} \frac{(\mathbf{A}_\lambda \tilde{U}_w)'(\mathbf{A}_\lambda \mathbf{D} \mathbf{A}_\lambda')^{-1}(\mathbf{A}_\lambda \tilde{U}_w)}{S} \sim F(u|(m-1)(h-1), g, \delta_\lambda) \end{cases} \quad (2.3.40)$$

em que $F(u|a, b, \delta)$ representa a distribuição F de Snedcor com a e b graus de liberdade e parâmetro de não centralidade δ .

Caso se detectem diferenças significativas para cultivares ou para anos, pode-se utilizar um ou o outro dos métodos de comparação múltipla já enunciados na subsecção 2.3.3. Neste caso os totais dos cultivares terão variâncias iguais a $\sum_{l=1}^h K_l$, enquanto que as variâncias dos totais anuais serão $mK_l, l = 1, \dots, h$. Assim, para o método de Scheffé, diferirão significativamente ao nível α , os cultivares cujos totais difiram por mais de

$$\Delta_\gamma = \sqrt{2(m-1) \left(\sum_{l=1}^h K_l \right) F_{1-\alpha, m-1, g} \frac{S}{g}}. \quad (2.3.41)$$

Quanto aos anos, diferirão os pares (l, l') cujos totais difiram por mais de

$$\Delta_{\eta, l, l'} = \sqrt{(h-1)m(K_l + K_{l'}) F_{1-\alpha, h-1, g} \frac{S}{g}}. \quad (2.3.42)$$

Como vimos na subsecção 2.3.3, se se quiser aplicar o método de Bonferroni, haverá que substituir nas expressões anteriores $\sqrt{(m-1)F_{1-\alpha, m-1, g}}$ e $\sqrt{(h-1)F_{1-\alpha, h-1, g}}$ pelos quantis da distribuição t de Student com g graus de liberdade para as probabilidades $1 - \frac{\alpha}{2(m-1)}$ e $1 - \frac{\alpha}{2(h-1)}$ respectivamente.

2.3.5 Testes de quociente de Verosimilhanças

Admitindo a normalidade e homocedasticidade do vector dos rendimentos \mathbf{y}_n , sendo C_i o conjunto dos índices dos cultivares presentes no i -ésimo bloco e n o número total de observações, ter-se-á, ver Pereira (2004), o logaritmo da verosimilhança

$$l(\boldsymbol{\gamma}_{2J}, \boldsymbol{\theta}_b, \sigma^2) = -\frac{1}{2\sigma^2} S(\boldsymbol{\gamma}_{2J}, \boldsymbol{\theta}_b) - \frac{n}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \sum_{i=1}^b \sum_{j \in C_i} \ln(p_{i,j}) \quad (2.3.43)$$

com $p_{i,j} = 0$ se $j \notin C_i$ e o estimador de máxima verosimilhança para σ^2 ,

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\gamma}_{2J}, \hat{\theta}_b). \quad (2.3.44)$$

$$\text{com } S(\gamma_{2J}, \theta_b) = \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} (y_{i,j} - \alpha_j - \beta_j \theta_i)^2.$$

Note-se que os estimadores de máxima verosimilhança e de mínimos quadrados coincidem quando se admite a normalidade de \mathbf{y}_n . Assim, os estimadores obtidos na secção anterior são estimadores de máxima verosimilhança. Observe-se que, a noção de verosimilhança, tem tido aplicações recentes de grande interesse, ver por exemplo Severine (2000). Vamos agora construir testes de quociente de verosimilhanças para hipóteses relativas aos coeficientes das regressões. Procuraremos completar a bateria de testes apresentados por Pereira (2004) e apresentaremos os mais importantes desses testes. Com $(\alpha_j, \beta_j), j = 1, \dots, J$ os pares de coeficientes para as regressões lineares a ajustar, testaremos em primeiro lugar a hipótese:

$$H_o : \alpha_1 = \dots = \alpha_J. \quad (2.3.45)$$

Seja α_0 o valor comum tomado pelos $\alpha_j, j = 1, \dots, J$, quando H_o se verifica. Sejam w o espaço paramétrico associado a H_o e Ω o espaço paramétrico total. Facilmente se verifica que $\dim(\Omega) - \dim(w) = J - 1$. Quando H_o se verifica, ver por exemplo, Mood et al (1974, pág. 441), $-2 \ln \lambda$, com

$$\lambda = \frac{\text{Sup}_w L(\alpha_J, \beta_J, \theta_b, \sigma^2)}{\text{Sup}_\Omega L(\alpha_J, \beta_J, \theta_b, \sigma^2)} \quad (2.3.46)$$

segue aproximadamente uma distribuição de qui-quadrado central com $r = J - 1$ graus de liberdade. Assim, ao nível de significância α , deve-se rejeitar H_o caso, $-2 \ln \lambda > \chi_{1-\alpha}^2(r)$, onde $\chi_{1-\alpha}^2(r)$ é o quantil para a probabilidade $1 - \alpha$ da distribuição do qui-quadrado central com r graus de liberdade. Para obter a estatística do teste, vamos calcular os supremos dos logaritmos da função de verosimilhança em Ω e em w .

Quando não se está a admitir H_o , o supremo do logaritmo da verosimilhança é

$$\text{Sup}_\Omega(l) = -\frac{n}{2} - \frac{n}{2} \ln \frac{S(\gamma_{2J}, \theta_b)}{n} - \frac{n}{2} \ln(2\pi) + \frac{1}{2} \sum_{i=1}^b \sum_{j \in C_i} \ln p_{i,j}. \quad (2.3.47)$$

Para obtermos o supremo associado a H_o , temos de maximizar o logaritmo da função de verosimilhança em w . Este logaritmo é dado por

$$\begin{aligned} l(\alpha_0 \mathbf{1}_J, \beta_J, \theta_b, \sigma^2) &= -\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} (y_{i,j} - \alpha_0 - \beta_j \theta_i)^2 + \frac{1}{2} \sum_{i=1}^b \sum_{j \in C_i} \ln p_{i,j} \\ &\quad - \frac{n}{2} \ln(2\pi\sigma^2) \end{aligned} \quad (2.3.48)$$

onde $\alpha_0 \mathbf{1}_J$ é um vector de dimensão J com todas as componentes iguais a α_0 . Veja-se que

$$\begin{aligned} \frac{\partial(l)}{\partial\sigma^2} = 0 &\Leftrightarrow \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} (y_{i,j} - \alpha_0 - \beta_j \theta_i)^2 - n\sigma^2 = 0 \\ &\Leftrightarrow \hat{\sigma}_w^2 = \frac{1}{n} \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} (y_{i,j} - \hat{\alpha}_{0,w} - \hat{\beta}_{j,w} \hat{\theta}_{i,w})^2 \end{aligned} \quad (2.3.49)$$

onde o índice inferior w , indica que são utilizados os estimadores obtidos maximizando a verosimilhança em w . Para calcular tais estimadores minimizamos

$$S_w(\alpha_0 \mathbf{1}_J, \boldsymbol{\beta}_J, \boldsymbol{\theta}_b) = \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} (y_{i,j} - \alpha_0 - \beta_j \theta_i)^2 \quad (2.3.50)$$

através do algoritmo Zig-Zag. Tomamos como ponto de partida $\boldsymbol{\theta}_{1,b} = \hat{\boldsymbol{\theta}}_b$, isto é o estimador de máxima verosimilhança para os índices ambientais obtido na ausência de restrições. As primeiras derivadas em ordem aos parâmetros α_0 e $\beta_j, j = 1, \dots, J$ de S_w são

$$\frac{\partial S_w}{\partial \alpha_0} = -2 \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} (y_{i,j} - \alpha_0 - \beta_j \theta_i) \quad (2.3.51)$$

$$\frac{\partial S_w}{\partial \beta_j} = -2 \sum_{i=1}^b p_{i,j} \theta_i (y_{i,j} - \alpha_0 - \beta_j \theta_i), j = 1, \dots, J. \quad (2.3.52)$$

Ora

$$\begin{aligned} \frac{\partial S_w}{\partial \beta_j} &= 0 \\ &\Leftrightarrow \sum_{i=1}^b p_{i,j} \theta_i (y_{i,j} - \alpha_0 - \beta_j \theta_i) = 0 \\ &\Leftrightarrow \sum_{i=1}^b p_{i,j} \theta_i y_{i,j} - \sum_{i=1}^b \alpha_0 p_{i,j} \theta_i - \sum_{i=1}^b p_{i,j} \beta_j \theta_i^2 = 0 \\ &\Leftrightarrow \alpha_0 = c_j - d_j \beta_j, j = 1, \dots, J \end{aligned} \quad (2.3.53)$$

com

$$c_j = \frac{\sum_{i=1}^b p_{i,j} \theta_i y_{i,j}}{\sum_{i=1}^b p_{i,j} \theta_i}, j = 1, \dots, J \quad (2.3.54)$$

e

$$d_j = \frac{\sum_{i=1}^b p_{i,j} \theta_i^2}{\sum_{i=1}^b p_{i,j} \theta_i}, j = 1, \dots, J. \quad (2.3.55)$$

Assim tem-se,

$c_1 - d_1\beta_1 = c_j - d_j\beta_j, j = 2, \dots, J$, logo obtém-se

$$\beta_j = e_j\beta_1 + f_j, j = 2, \dots, J \quad (2.3.56)$$

com $e_j = \frac{d_1}{d_j}$ e $f_j = \frac{c_j - c_1}{d_j}, j = 2, \dots, J$.

Resolvendo $\sum_{j \in C_i} \frac{\partial S_w}{\partial \beta_j} = 0$ em ordem a α_0 , vem:

$$-\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i y_{i,j} + \left(\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i \right) \alpha_0 + \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i^2 \beta_j = 0 \quad (2.3.57)$$

obtendo-se:

$$\alpha_0 = \frac{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i y_{i,j} - \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i^2 (e_j \beta_1 + f_j)}{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i} \quad (2.3.58)$$

com $e_1 = 1$ e $f_1 = 0$.

Por outro lado, resolvendo $\frac{\partial S_w}{\partial \alpha_0} = 0$ em ordem a α_0 , obtém-se:

$$\alpha_0 = \frac{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} y_{i,j} - \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i (e_j \beta_1 + f_j)}{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j}}. \quad (2.3.59)$$

Igualando os segundos membros das expressões (2.3.58) e (2.3.59) e resolvendo em ordem a β_1 obtém-se,

$$\beta_1 = \frac{C_1 - C_2}{D_1 - D_2} \quad (2.3.60)$$

com

$$C_1 = \frac{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} y_{i,j} \theta_i - \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i^2 f_j}{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i} \quad (2.3.61)$$

$$C_2 = \frac{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} y_{i,j} - \sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i f_j}{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j}} \quad (2.3.62)$$

$$D_1 = \frac{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i^2 e_j}{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i} \quad (2.3.63)$$

$$D_2 = \frac{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j} \theta_i e_j}{\sum_{i=1}^b \sum_{j \in C_i} p_{i,j}}. \quad (2.3.64)$$

São agora fáceis de obter os valores de α_0 e de $\beta_j = \frac{c_j - \alpha}{d_j}$, $j = 2, \dots, J$, correspondentes à primeira iteração. Continuando a aplicar o algoritmo Zig-Zag, vamos minimizar $\sum_{i=1}^b h_i(\theta_i)$ onde

$$h_i(\theta_i) = \sum_{j \in C_i} p_{i,j} (y_{i,j} - \hat{\alpha}_0 - \hat{\beta}_j \theta_i)^2, i = 1, \dots, b, \quad (2.3.65)$$

vindo

$$\theta_i = \frac{\sum_{j \in C_i} p_{i,j} (y_{i,j} - \hat{\alpha}_0) \hat{\beta}_j}{\sum_{j \in C_i} p_{i,j} \hat{\beta}_j^2}, i = 1, \dots, b. \quad (2.3.66)$$

Pode proceder-se à estandardização, por forma a manter inalterados o valor máximo e mínimo dos índices ambientais e repetir o processo, até que o valor de S_w estabilize.

Designa-se por S_w^* o valor final obtido para S_w . Pode-se agora calcular $\hat{\sigma}_w^2 = \frac{S_w^*}{n}$ e o supremo do logaritmo da função de verosimilhança em w :

$$Sup_w(l) = -\frac{n}{2} - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \frac{S_w^*}{n} + \frac{1}{2} \sum_{i=1}^b \sum_{j \in C_i} \ln p_{i,j}. \quad (2.3.67)$$

Por fim, calcula-se a estatística de teste e realiza-se o ensaio da hipótese H_o . Este teste foi inicialmente obtido por Pereira (2004), tendo o mesmo autor mostrado como testar a hipótese de igualdade dos β_1, \dots, β_J , e ainda a hipótese de igualdade simultânea dos $\alpha_1, \dots, \alpha_J$ e dos β_1, \dots, β_J .

Consideremos agora um teste para testar a hipótese

$$H_o : \beta_j = \beta_{j'} = \beta, j \neq j', j, j' = 1, \dots, J. \quad (2.3.68)$$

Verificamos que o espaço paramétrico total Ω , mantém-se inalterado relativamente ao utilizado no teste anterior, ao passo que agora se tem,

$$\dim(\Omega) - \dim(w) = 1. \quad (2.3.69)$$

Assim, quando H_o se verifica, tem-se $-2\ln(\lambda) \sim \chi_1^2$, com

$$\lambda = \frac{\text{Sup}_w L(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J, \boldsymbol{\theta}_b, \sigma^2)}{\text{Sup}_\Omega L(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J, \boldsymbol{\theta}_b, \sigma^2)}. \quad (2.3.70)$$

Para obter a estatística do teste temos de, tal como no caso anterior, calcular os supremos do logaritmo da função de verosimilhança em Ω e w . O supremo em Ω é o mesmo que para o teste anterior. Temos agora que proceder ao cálculo do supremo em w . Dada a simetria do problema, tomamos $j = 1$ e $j' = 2$ o que nos leva a minimizar em ordem aos $\alpha_1, \dots, \alpha_J$ e aos $\beta_1, \beta_3, \dots, \beta_J$ a função,

$$\begin{aligned} S_w &= \sum_{i=1}^b \left[\sum_{j=1}^2 p_{i,j} (y_{i,j} - \alpha_j - \beta\theta_i)^2 + \sum_{j=3}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j\theta_i)^2 \right] \\ &= S_{w,1} + S_{w,2}, \end{aligned} \quad (2.3.71)$$

com

$$\begin{cases} S_{w,1} = \sum_{i=1}^b \sum_{j=1}^2 p_{i,j} (y_{i,j} - \alpha_j - \beta\theta_i)^2 \\ S_{w,2} = \sum_{i=1}^b \sum_{j=3}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j\theta_i)^2 \end{cases} \quad (2.3.72)$$

Haverá pois que minimizar $S_{w,1}$ em ordem a α_1, α_2 e β e $S_{w,2}$ em ordem a $(\alpha_j, \beta_j), j = 3, \dots, J$. A segunda minimização dá origem aos estimadores dos mínimos quadrados. Temos pois que considerar a minimização de $S_{w,1}$, para a qual se obtêm as derivadas

$$\begin{aligned} \frac{\partial S_{w,1}}{\partial \alpha_j} &= -2 \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta\theta_i), j = 1, 2 \\ \frac{\partial S_{w,1}}{\partial \beta} &= -2 \sum_{i=1}^b \sum_{j=1}^2 p_{i,j} \theta_i (y_{i,j} - \alpha_j - \beta\theta_i), \end{aligned} \quad (2.3.73)$$

cujos anulamentos nos conduzem às equações que integram o sistema,

$$\begin{cases} -2 \sum_{i=1}^b p_{i,1} (y_{i,1} - \alpha_1 - \beta\theta_i) = 0 \\ -2 \sum_{i=1}^b p_{i,2} (y_{i,2} - \alpha_2 - \beta\theta_i) = 0 \\ -2 \sum_{i=1}^b [p_{i,1} \theta_i (y_{i,1} - \alpha_1 - \beta\theta_i) + p_{i,2} \theta_i (y_{i,2} - \alpha_2 - \beta\theta_i)] = 0 \end{cases} \quad (2.3.74)$$

A solução deste sistema dar-nos-á os valores de α_1 , α_2 e β . Da primeira equação obtém-se,

$$\alpha_1 = \frac{\sum_{i=1}^b p_{i,1} y_{i,1} - \sum_{i=1}^b p_{i,1} \beta \theta_i}{\sum_{i=1}^b p_{i,1}} \quad (2.3.75)$$

e da segunda

$$\alpha_2 = \frac{\sum_{i=1}^b p_{i,2} y_{i,2} - \sum_{i=1}^b p_{i,2} \beta \theta_i}{\sum_{i=1}^b p_{i,2}}. \quad (2.3.76)$$

Substituindo α_1 e α_2 na terceira equação obtém-se,

$$\beta = \frac{A_1 - A_2 - A_3}{c} \quad (2.3.77)$$

com

$$\left\{ \begin{array}{l} A_1 = \sum_{i=1}^b \sum_{j=1}^2 p_{i,j} y_{i,j} \theta_i \\ A_2 = \sum_{i=1}^b p_{i,1} \alpha_1 \theta_i \\ A_3 = \sum_{i=1}^b p_{i,2} \alpha_2 \theta_i \\ c = \sum_{i=1}^b \theta_i^2 (p_{i,1} + p_{i,2}) \end{array} \right. \quad (2.3.78)$$

Temos agora todos os valores estimados, para os parâmetros correspondentes à primeira fase da primeira iteração. Podemos aplicar o algoritmo Zig-Zag a partir destes valores, procedendo à minimização das funções $\sum_{i=1}^b h_i(\theta_i)$, com

$$h_i(\theta_i) = \sum_{j=1}^2 p_{i,j} (y_{i,j} - \alpha_j - \beta \theta_i)^2 + \sum_{j=3}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j \theta_i)^2, i = 1, \dots, b. \quad (2.3.79)$$

Fazendo $\frac{\partial h_i(\theta_i)}{\partial \theta_i} = 0$ e resolvendo em ordem a θ_i , obtém-se então:

$$\theta_i = \frac{\sum_{j=1}^2 p_{i,j} \beta (y_{i,j} - \alpha_j) + \sum_{j=3}^J p_{i,j} \beta_j (y_{i,j} - \alpha_j)}{\beta^2 \sum_{j=1}^2 p_{i,j} + \sum_{j=3}^J p_{i,j} \beta_j^2}, i = 1, \dots, b. \quad (2.3.80)$$

À semelhança do teste anterior, podemos standardizar os índices ambientais e repetir as iterações até que o valor de S_w estabilize. Sendo S_w^* o valor final obtido, determina-se $\hat{\sigma}_w^2 = \frac{S_w^*}{n}$ e obtemos então o valor do supremo do logaritmo da função de verosimilhança em w ,

$$Sup_w(l) = -\frac{n}{2} - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \frac{S_w^*}{n} + \frac{1}{2} \sum_{i=1}^b \sum_{j \in C_i} \ln p_{i,j}. \quad (2.3.81)$$

Finalmente, está-se em condições de calcular a estatística de teste e realizar o ensaio da hipótese H_o . Este teste é o primeiro duma série interessante. Com efeito, poderemos querer justificar um agrupamento de cultivares em L grupos com s_1, \dots, s_L cultivares. Para tal, após a reordenação dos cultivares dão-se índices contíguos aos cultivares do mesmo grupo. Com $\bar{s}_0 = 0$, $\bar{s}_l = \sum_{n=1}^l s_n, l = 1, \dots, L$, os índices dos cultivares do l -ésimo grupo serão $\bar{s}_{l-1} + 1, \dots, \bar{s}_l$. Neste caso passamos a ter apenas $2L$ coeficientes distintos quando a hipótese se verifica, vindo, $dim(\Omega) - dim(w) = 2(J - L)$ e além disso $S_w = \sum_{l=1}^L S_{w,l}$ com

$$S_{w,l} = \sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} (y_{i,j} - \hat{\alpha}_l - \hat{\beta}_l \theta_i)^2, l = 1, \dots, L, \quad (2.3.82)$$

onde $(\hat{\alpha}_l, \hat{\beta}_l)$ são valores supostos iguais para os coeficientes das regressões do l -ésimo grupo.

Ora o sistema

$$\begin{cases} \frac{\partial S_{w,l}}{\partial \hat{\alpha}_l} = -2 \sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} (y_{i,j} - \hat{\alpha}_l - \hat{\beta}_l \theta_i) = 0; l = 1, \dots, L \\ \frac{\partial S_{w,l}}{\partial \hat{\beta}_l} = -2 \sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} (y_{i,j} - \hat{\alpha}_l - \hat{\beta}_l \theta_i) \theta_i = 0; l = 1, \dots, L \end{cases} \quad (2.3.83)$$

tem a solução

$$\tilde{\beta}_l = \frac{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} (y_{i,j} - y_{o,j}) (\theta_i - \theta_{o,l})}{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} (\theta_i - \theta_{o,l})^2}, l = 1, \dots, L \quad (2.3.84)$$

e

$$\tilde{\alpha}_l = \frac{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} y_{i,j} - \tilde{\beta}_l \sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} \theta_i}{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j}}, l = 1, \dots, L \quad (2.3.85)$$

sendo

$$\theta_{o,l} = \frac{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} \theta_i}{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j}}, l = 1, \dots, L \quad (2.3.86)$$

e

$$y_{o,j} = \frac{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j} y_{i,j}}{\sum_{j=\bar{s}_{l-1}+1}^{\bar{s}_l} \sum_{i=1}^b p_{i,j}}, j = 1, \dots, J. \quad (2.3.87)$$

Para obter os valores destas expressões basta interpretar $S_{w,l}$ como a soma de quadrados que se minimiza para ajustar uma única regressão linear aos dados dos cultivares do grupo $l, l = 1, \dots, L$. Obtém-se assim o supremo, $S_w^* = \sum_{l=1}^L S_{w,l}^*$ debaixo da hipótese testada e constrói-se o teste. O número de graus de liberdade será $r = 2J - 2L = 2(J - L)$.

2.3.6 Regiões de confiança

Vamos agora utilizar uma outra abordagem possível, CANE (completely asymptotical normal estimators), ver Mexia (1992). Para simplificar a notação vamos, nesta secção, omitir os índices inferiores dos vectores.

Sendo $\tilde{\gamma}$ e $\tilde{\theta}$ os vectores dos coeficientes de regressão e índices ambientais ajustados respectivamente, admitamos que

$$\tilde{\gamma} \sim N(\gamma, \sigma^2(\mathbf{X}'(\boldsymbol{\theta})\mathbf{X}(\boldsymbol{\theta}))^{-1}), \quad (2.3.88)$$

isto é que $\tilde{\gamma}$ é normal com vector médio γ e matriz de covariância $\sigma^2(\mathbf{X}'(\boldsymbol{\theta})\mathbf{X}(\boldsymbol{\theta}))^{-1}$ independente de

$$S = S(\tilde{\gamma}, \tilde{\theta}), \quad (2.3.89)$$

que se distribuirá como o produto de σ^2 por um qui-quadrado central com $g = n - \text{Car}(\mathbf{X}(\tilde{\theta}))$ graus de liberdade, escrevendo-se $S \sim \sigma^2 \chi_g^2$. Sendo $x_{g,1-\alpha}$ o quantil para a probabilidade $1 - \alpha$ de um χ_g^2 , quando $\mathbf{X}(\tilde{\theta}) \approx \mathbf{X}(\boldsymbol{\theta})$ ter-se-á, ver Mexia(1992),

$$\text{prob}\{(\gamma - \tilde{\gamma})'[\mathbf{X}'(\tilde{\theta})\mathbf{X}(\tilde{\theta})](\gamma - \tilde{\gamma}) \leq x_{g,1-\alpha} \frac{S}{g}\} \approx 1 - \alpha, \quad (2.3.90)$$

obtendo-se assim um elipsóide de confiança de nível aproximado a $1 - \alpha$ para γ .

Regressando ao algoritmo Zig-Zag, em cada iteração o vector dos índices ambientais é estimado através da expressão (2.2.17), que é da forma,

$$\boldsymbol{\theta} = \mathbf{A}(\boldsymbol{\gamma})\mathbf{y} - \mathbf{b}(\boldsymbol{\gamma}), \quad (2.3.91)$$

onde \mathbf{A} e \mathbf{B} são respectivamente a matriz e o vector convenientes, ambos função de $\boldsymbol{\gamma}$. Admitindo-se que, $\mathbf{A}(\tilde{\boldsymbol{\gamma}}) \approx \mathbf{A}(\boldsymbol{\gamma})$ e $\mathbf{b}(\tilde{\boldsymbol{\gamma}}) \approx \mathbf{b}(\boldsymbol{\gamma})$, a mesma perspectiva permite-nos passar a admitir que, $\tilde{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2 \mathbf{A}(\tilde{\boldsymbol{\gamma}}) \mathbf{A}'(\tilde{\boldsymbol{\gamma}}))$, obtendo-se agora

$$prob\{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' [\mathbf{A}(\tilde{\boldsymbol{\gamma}}) \mathbf{A}'(\tilde{\boldsymbol{\gamma}})]^{-1} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \leq x_{g,1-\alpha} \frac{S}{g}\} \approx 1 - \alpha \quad (2.3.92)$$

isto é, um elipsóide de confiança de nível aproximado $1 - \alpha$ para $\boldsymbol{\theta}$. Abriu-se assim uma possibilidade para desenvolvimentos futuros. Não a explorámos devido ao nosso objectivo central ser a aplicação da ACR a planos de melhoramento.

2.4 Classificação e Selecção de Cultivares

Nesta secção vamos abordar o problema da classificação e selecção de cultivares, utilizando para isso testes sobre hipóteses relativas aos coeficientes de regressão. Para a classificação apresentamos nesta secção os testes a utilizar, para a selecção vamos utilizar os testes apresentados na subsecção 2.3.3 em conjunto com o método do contorno superior. Testaremos assim hipóteses sobre funções lineares e aplicaremos o teste t bem como os métodos de comparação múltipla de Scheffé e Bonferroni.

2.4.1 Classificação

Sejam $(\tilde{\alpha}_j, \tilde{\beta}_j)$, $j = 1, \dots, J$ os valores estimados dos coeficientes de regressão para os cultivares utilizados. Os cultivares serão classificados em intensivos, extensivos e estáveis. Os primeiros serão aqueles cuja inclinação da respectiva recta de regressão ajustada é significativamente superior a 1, ao nível de 5%, os segundos aqueles cuja inclinação da recta de regressão ajustada é significativamente inferior a 1, ao mesmo nível e os restantes serão os cultivares estáveis. Somos assim levados a testar as hipóteses

$$H_{o,\beta_j} : \beta_j = 1, j = 1, \dots, J \quad (2.4.1)$$

contra

$$H_{1,\beta_j} : \beta_j > 1, j = 1, \dots, J \quad (2.4.2)$$

para os cultivares tais que $\tilde{\beta}_j > 1, j = 1, \dots, J$ e

$$H_{o,\beta_j} : \beta_j = 1, j = 1, \dots, J \quad (2.4.3)$$

contra

$$H_{2,\beta_j} : \beta_j < 1, j = 1, \dots, J \quad (2.4.4)$$

para os cultivares tais que $\tilde{\beta}_j < 1, j = 1, \dots, J$. Para testar estas hipóteses, utilizamos a estatística de teste

$$t = \frac{\tilde{\beta}_j - \beta_j}{\sqrt{\frac{1}{s_j} \frac{S}{g}}} \quad (2.4.5)$$

que, quando a hipótese H_{o,β_j} se verifica, tem distribuição t de Student com g graus de liberdade. Com os resultados obtidos nestes testes, podemos classificar os cultivares nos quatro cenários definidos na tabela 2.1,

Tabela 2.1: Grelha de classificação dos cultivares

	com significância	sem significância
$\tilde{\beta} > 1$	cenário I	cenário II
$\tilde{\beta} < 1$	cenário III	cenário IV

Os cultivares que se enquadram no cenário *I* dizem-se significativamente intensivos e os que se enquadram no cenário *III* dizem-se significativamente extensivos. Os cultivares que se enquadram nos cenários *II* e *IV* dizem-se estáveis.

Consideraremos ainda as ordenadas na origem para o modelo modificado

$$y'_j = \alpha'_j + \beta'_j(x - \theta_o), j = 1, \dots, J, \quad (2.4.6)$$

com θ_o o ponto médio do intervalo de variação dos índices ambientais. Tem interesse testar a hipótese de nulidade dos $\alpha'_j, j = 1, \dots, J$. Os estimadores de que dispomos correspondem aos ajustamentos efectuados com base no modelo dado pela subsecção 2.2.2, logo não necessitamos de fazer novos ajustamentos, temos apenas de relacionar os estimadores correspondentes nos dois modelos, vendo-se que,

$$\alpha'_j = \alpha_j + \beta_j \theta_o, j = 1, \dots, J, \quad (2.4.7)$$

o que nos permite tratar os α'_j como funções lineares. Se α'_j for significativamente positivo, o cultivar com índice j terá efeito global positivo, enquanto que se for significativamente negativo, o cultivar com índice j terá efeito global negativo.

Assim, testaremos as hipóteses

$$H_{o,\alpha'_j} : \alpha'_j = 0, j = 1, \dots, J \quad (2.4.8)$$

contra

$$H_{1,\alpha'_j} : \alpha'_j > 0, j = 1, \dots, J \quad (2.4.9)$$

para os cultivares tais que $\tilde{\alpha}'_j > 0, j = 1, \dots, J$ e a hipótese

$$H_{o,\alpha'_j} : \alpha'_j = 0, j = 1, \dots, J \quad (2.4.10)$$

contra

$$H_{2,\alpha'_j} : \alpha'_j < 0, j = 1, \dots, J \quad (2.4.11)$$

para os cultivares tais que $\tilde{\alpha}'_j < 0, j = 1, \dots, J$. Utilizamos para testar estas hipóteses a estatística

$$t = \frac{\tilde{\alpha}'_j - \alpha'_j}{\sqrt{Var(\tilde{\alpha}'_j)}}, \quad (2.4.12)$$

onde

$$Var(\tilde{\alpha}'_j) = \frac{S}{g} \left[\frac{1}{b_j} + \frac{x_{.j}^2}{s_j} + \frac{\theta_o^2}{s_j} - 2\theta_o \frac{x_{.j}}{s_j} \right] \quad (2.4.13)$$

que, quando H_{o,α'_j} se verifica, tem distribuição t de Student com g graus de liberdade.

Serão claramente preferidos os cultivares cujo valor de $\tilde{\alpha}'_j$ seja significativamente superior a zero. Os mesmos deverão ser ordenados por ordem decrescente, por forma a facilitar a identificação daqueles que têm um valor mais elevado. Esses cultivares apresentam rendimentos significativos mesmo quando o índice ambiental toma valores mínimos.

Para ambientes de alta performance, devemos seleccionar cultivares que se enquadrem no cenário I , com $\tilde{\alpha}'$ elevado, comparativamente com os demais cultivares. Para ambientes de média performance, devem escolher-se cultivares que se enquadrem num dos cenários II ou IV com $\tilde{\alpha}'$ preferencialmente elevado. Resta-nos dizer que, para ambientes de baixa performance, são mais adequados os cultivares extensivos.

2.4.2 Selecção pelo método do Contorno Superior

A utilização do contorno superior no processo de selecção de cultivares foi introduzida por Mexia et al (1997). A informação fornecida pelo traçado do contorno superior, permite facilmente identificar um conjunto de cultivares, que se destacam do grupo pelo facto de apresentarem produções máximas, para certos valores do índice ambiental. Sejam

$$\tilde{y}_j = \tilde{\alpha}_j + \tilde{\beta}_j \theta, j = 1, \dots, J \quad (2.4.14)$$

as J regressões ajustadas, uma por cultivar e $\tilde{\theta}_b = [\tilde{\theta}_1, \dots, \tilde{\theta}_b]'$ o vector dos índices ambientais ajustado e estandardizado, ponhamos

$$\begin{cases} \theta_{max} = \max\{\tilde{\theta}_1, \dots, \tilde{\theta}_b\} \\ \theta_{min} = \min\{\tilde{\theta}_1, \dots, \tilde{\theta}_b\} \end{cases} \quad (2.4.15)$$

designando por regressões dominantes, no intervalo $[\theta_{min}, \theta_{max}]$ de variação de θ , aquelas a que correspondem as produções ajustadas máximas. Obviamente os cultivares associados às regressões dominantes serão os cultivares dominantes, para os correspondentes intervalos de variação dos índices ambientais, designados intervalos de dominância. A selecção deve dar prioridade aos cultivares que integram o contorno superior e proceder a comparações entre esses cultivares dominantes e todos os restantes, dos quais alguns serão significativamente dominados, devendo ser eliminados do processo de selecção. Será muito interessante obtermos cultivares dominantes, cujo valor de $\tilde{\alpha}'$ seja elevado e que segundo a tabela 2.1 de classificação de cultivares se enquadrem no cenário *I*.

Geometria

Antes de prosseguir, vamos atribuir índices às regressões ajustadas, ordenando-as por ordem decrescente do seu declive. Assim, à regressão com maior declive, será atribuído o índice 1 e à regressão menos inclinada será atribuído o índice J ou seja,

$$\tilde{\beta}_1 \geq \tilde{\beta}_2 \geq \dots \geq \tilde{\beta}_J \quad (2.4.16)$$

tendo-se $\tilde{\alpha}_j > \tilde{\alpha}_{j+1}$ se $\tilde{\beta}_j = \tilde{\beta}_{j+1}$.

Em seguida, iremos estabelecer alguns resultados que permitem caracterizar geometricamente o traçado do contorno superior. Começemos por estabelecer a

Proposição 2.9 *O contorno superior é uma poligonal convexa.*

Dem: O contorno superior é uma poligonal constituída por segmentos pertencentes às regressões dominantes. Quando duas rectas se intersectam, à direita do ponto de intersecção, fica por cima a recta que tem maior declive, e à esquerda a que tem menor declive. Sendo assim, verifica-se que os coeficientes angulares dos troços que constituem o contorno superior, crescem da esquerda para a direita o que completa a demonstração da proposição. ■

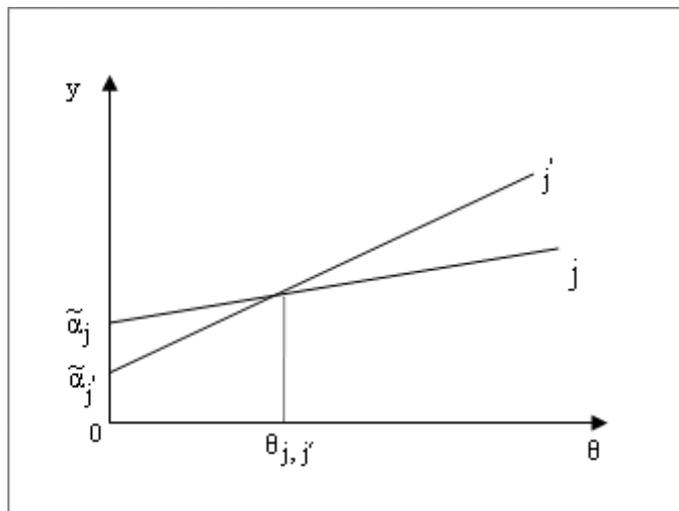
Quando se tem $\tilde{\beta}_j \neq \tilde{\beta}_{j'}$

$$\theta_{j,j'} = \frac{\tilde{\alpha}_j - \tilde{\alpha}_{j'}}{\tilde{\beta}_{j'} - \tilde{\beta}_j}, \quad j \neq j' \text{ e } j, j' = 1, \dots, J \quad (2.4.17)$$

é o ponto de intersecção entre as rectas de índices j e j' . Se não existir nenhum ponto de intersecção entre duas rectas com índices j e j' , as mesmas serão paralelas e integrará o

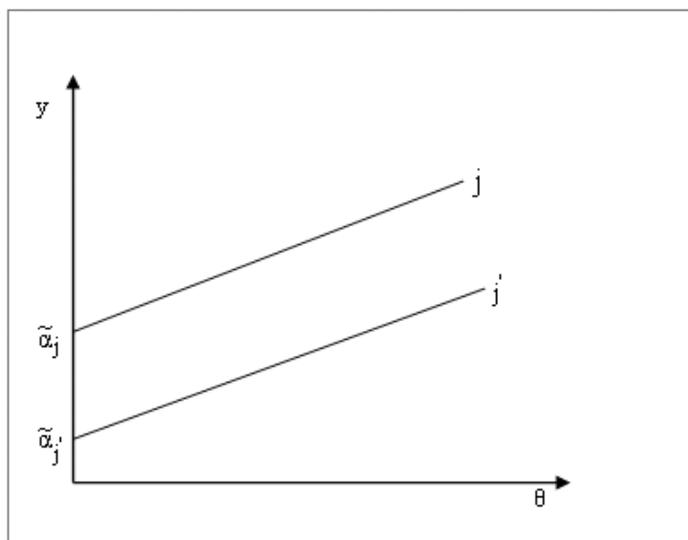
contorno superior a recta com maior ordenada na origem. Na figura 2.1 ilustra-se a situação onde se tem $\tilde{\beta}_j < \tilde{\beta}_{j'}$ e portanto $j > j'$.

Figura 2.1: Ponto de Intersecção



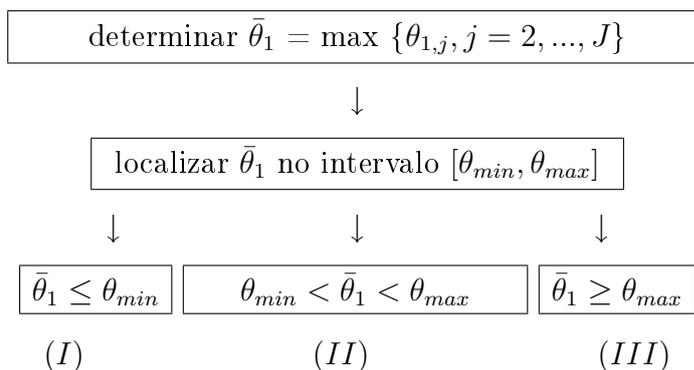
Na figura 2.2 ilustra-se o caso em que as duas rectas são paralelas, sendo obviamente a recta de índice j a que integra o contorno superior.

Figura 2.2: Rectas Paralelas



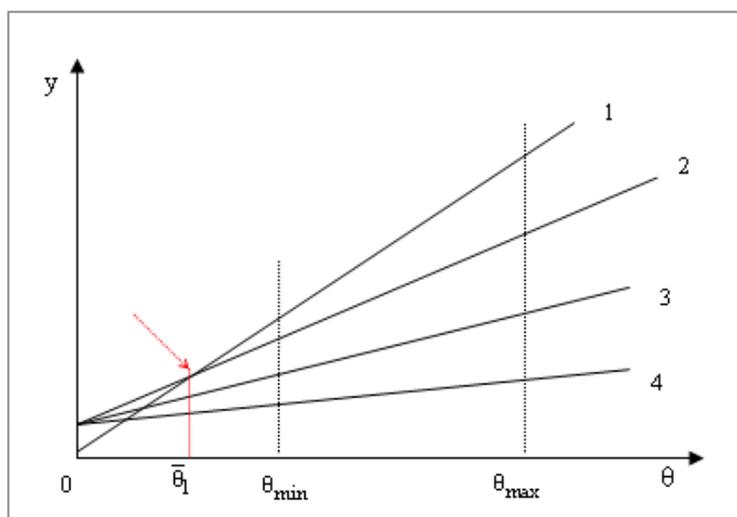
Para construir o contorno superior, pode-se começar por seguir o Esquema 2.

Esquema 2 - Determinação do Contorno Superior



No Caso *I* o contorno superior reduz-se à recta ajustada com índice 1, como se pode constatar no exemplo gráfico ilustrado na figura 2.3, onde temos $J = 4$.

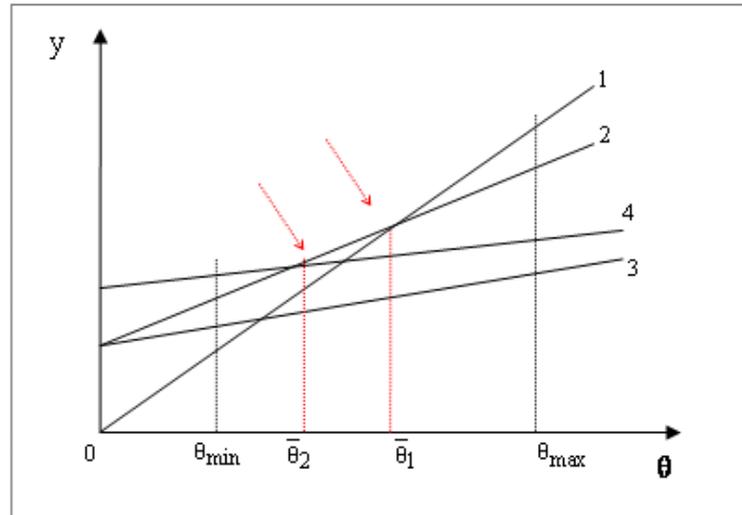
Figura 2.3: Traçado do contorno superior quando $\bar{\theta}_1 \leq \theta_{min}$



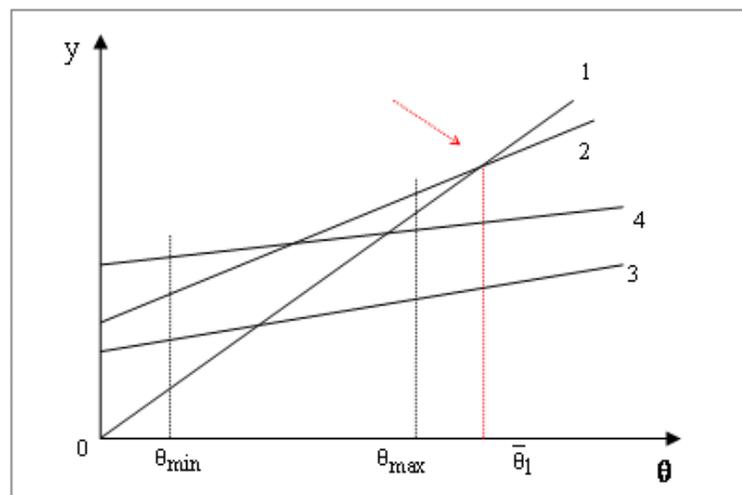
Quanto ao Caso *II*, tem-se $\theta_{min} < \bar{\theta}_1 < \theta_{max}$ sendo o cultivar com índice 1 dominante em $[\bar{\theta}_1, \theta_{max}]$. Seja j o índice do cultivar cuja recta de regressão intersecta a do cultivar com índice 1 no ponto de abcissa $\bar{\theta}_1$. No intervalo $[\theta_{min}, \bar{\theta}_1]$ o cultivar com índice j é dominante, podendo existir mais cultivares dominantes para além deste, partilhando nesse caso o intervalo referido. Para averiguar sobre a existência de mais cultivares dominantes terá de se obter

$$\bar{\theta}_2 = \max \{ \theta_{j,j'}, j' = j + 1, \dots, J \}. \quad (2.4.18)$$

Para o domínio $[\theta_{min}, \bar{\theta}_2]$, terá de se proceder do mesmo modo por forma a obter $\bar{\theta}_3$ e assim sucessivamente, até que se tenha o contorno superior definido entre θ_{min} e θ_{max} . Com $J = 4$ cultivares vejamos o exemplo ilustrado pela figura 2.4.

Figura 2.4: Traçado do contorno superior quando $\theta_{min} < \bar{\theta}_1 < \theta_{max}$ 

No Caso *III* o facto de se ter $\bar{\theta}_1 \geq \theta_{max}$, implica que a recta de regressão com índice 1 não fará parte do contorno superior e portanto o cultivar associado nunca será dominante. Ter-se-á que trabalhar com o conjunto de rectas com índices $j = 2, \dots, J$ e determinar $\bar{\theta}_2 = \max\{\theta_{2,j}, j = 3, \dots, J\}$. Caso $\bar{\theta}_2$ se encontre no intervalo $[\theta_{min}, \theta_{max}]$, estar-se-á numa situação idêntica à definida pelo Caso *II*. Se ainda se verificar que $\bar{\theta}_2 \geq \theta_{max}$, o cultivar com índice 2 não será dominante passando a trabalhar-se apenas com o conjunto de rectas de regressão com índices $j' = 3, \dots, J$ e assim sucessivamente até estarmos na situação definida pelos Casos *II*, ou *I*. O Caso *III* pode-se ilustrar pelo exemplo dado na figura 2.5.

Figura 2.5: Traçado do contorno superior quando $\bar{\theta}_1 \geq \theta_{max}$ 

Note-se que as três últimas figuras apresentadas são apenas exemplos que fazem parte de um conjunto muito vasto de diferentes casos possíveis.

O caso *II* merece uma discussão mais elaborada, pela diversidade de possibilidades que o integram. Vamos assim, para este caso, estabelecer alguns resultados que caracterizam o traçado do contorno superior no intervalo $[\theta_{min}, \bar{\theta}_1]$. Suponhamos que $\bar{\theta}_1 = \theta_{1,j'}$, interessando-nos apenas agora os cultivares com índices $j = j', \dots, J$. Começemos por estabelecer

Proposição 2.10 *Se $\bar{\theta}_1 = \theta_{1,j'}$ e para $j \neq 1, j', \theta_{1,j'} > \theta_{1,j}$ o cultivar j' é dominante.*

Dem: Para quaisquer $j \neq 1, j'$ e para $\theta_{1,j} < \theta < \theta_{1,j'}$ tem-se a desigualdade $\tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta > \tilde{\alpha}_1 + \tilde{\beta}_1\theta > \tilde{\alpha}_j + \tilde{\beta}_j\theta$ e portanto o cultivar com índice j' é dominante. ■

Suponhamos que em $\theta = \bar{\theta}_1$ intersectam a recta de índice 1 mais do que uma recta de regressão. Seja C o conjunto dos índices dessas rectas e j' o maior índice do conjunto C . Estabeleçamos a

Proposição 2.11 *Se j' é o maior índice para o qual $\bar{\theta}_1 = \theta_{1,j'}$, o cultivar com índice j' é dominante.*

Dem: Esta proposição estabelece que, das rectas que intersectam a primeira, em $\bar{\theta}_1$, irá pertencer ao contorno superior aquela que tem menor inclinação. É fácil de ver que, devido à intersecção das rectas com índices 1 e j' , temos $\tilde{\beta}_{j'} < \tilde{\beta}_1$. Se tivermos $1 < j < j'$, virá $\tilde{\beta}_{j'} < \tilde{\beta}_j < \tilde{\beta}_1$, $j = 2, \dots, j' - 1$. Ora, se $j \in C - \{j'\}$, para $\theta > \bar{\theta}_1$, vem $\tilde{\alpha}_1 + \tilde{\beta}_1\theta > \tilde{\alpha}_j + \tilde{\beta}_j\theta > \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta$, verificando-se a desigualdade contrária se $\theta < \bar{\theta}_1$. Caso $j \notin C$, temos $\theta_{1,j} < \theta_{1,j'}$ e para qualquer $\theta \in]\theta_{1,j}, \theta_{1,j'}[$ tem-se a desigualdade $\tilde{\alpha}_j + \tilde{\beta}_j\theta < \tilde{\alpha}_1 + \tilde{\beta}_1\theta < \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta$. Portanto, à esquerda de $\bar{\theta}_1$, existe um intervalo de valores do índice ambiental onde o cultivar j' é dominante. ■

Vejamos agora outro caso através da

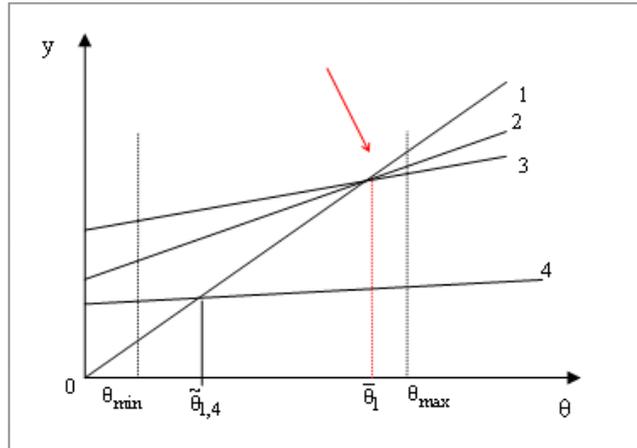
Proposição 2.12 *Quando $1 < j < j'$ o cultivar com índice j não é dominante.*

Dem: Tem-se $\tilde{\beta}_{j'} < \tilde{\beta}_j < \tilde{\beta}_1$ bem como

$\tilde{\alpha}_j + \tilde{\beta}_j\bar{\theta}_1 \leq \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\bar{\theta}_1 = \tilde{\alpha}_1 + \tilde{\beta}_1\bar{\theta}_1$, logo quando $\theta < \bar{\theta}_1$, ter-se-á a desigualdade $\tilde{\alpha}_j + \tilde{\beta}_j\theta < \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta$ e quando $\theta > \bar{\theta}_1$, ter-se-á a desigualdade $\tilde{\alpha}_j + \tilde{\beta}_j\theta < \tilde{\alpha}_1 + \tilde{\beta}_1\theta$, vendo-se facilmente que em nenhum destes casos o cultivar com índice j é dominante, o que estabelece a tese. ■

A figura 2.6 ilustra um exemplo da situação prevista pela última proposição, onde com $J = 4$ temos dois cultivares dominantes.

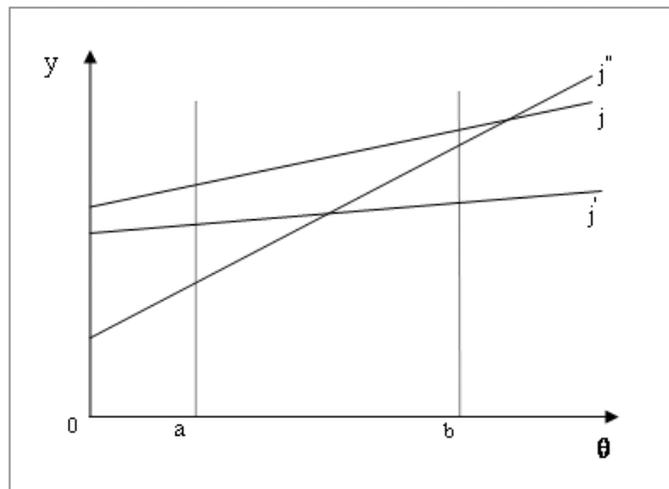
Figura 2.6: Contorno superior com dois cultivares dominantes



Comparação de cultivares

Suponha-se a situação ilustrada pela figura 2.7, onde se têm os cultivares j , j' e j'' , sendo j o índice do cultivar dominante no intervalo $[a, b]$.

Figura 2.7: Comparação entre cultivares dominantes e não dominantes



O cultivar j' terá de ser comparado com o cultivar j , para $\theta = a$ pois é para esses valores que a diferença $\tilde{\alpha}_j + \tilde{\beta}_j\theta - (\tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta)$ é mínima no intervalo $[a, b]$. Por sua vez, a comparação entre os cultivares j e j'' , deve ser efectuada para $\theta = b$, onde a diferença $\tilde{\alpha}_j + \tilde{\beta}_j\theta - (\tilde{\alpha}_{j''} + \tilde{\beta}_{j''}\theta)$ é mínima. Para proceder a tais comparações, consideramos as hipóteses:

$$H_{o,j,j'} : \alpha_j + \beta_j a = \alpha_{j'} + \beta_{j'} a \text{ contra } H_{1,j,j'} : \alpha_j + \beta_j a > \alpha_{j'} + \beta_{j'} a$$

e

$H_{0,j,j''} : \alpha_j + \beta_j b = \alpha_{j''} + \beta_{j''} b$ contra $H_{1,j,j''} : \alpha_j + \beta_j b > \alpha_{j''} + \beta_{j''} b$. Estes pares de hipóteses são casos particulares de

$$\begin{cases} H_{0,j,j'}(1, \theta_o) : U_j(1, \theta_o) = U_{j'}(1, \theta_o) \\ H_{1,j,j'}(1, \theta_o) : U_j(1, \theta_o) > U_{j'}(1, \theta_o) \end{cases} \quad (2.4.19)$$

onde se tem

$U_j(1, \theta_o) = \alpha_j + \beta_j \theta_o, j = 1, \dots, J; j' = 1, \dots, J, j \neq j'$. Para testar estas hipóteses utilizamos a estatística de teste dada por (2.3.16) da subsecção 2.3.3. Quando necessário podemos ainda proceder à realização dos testes de comparação múltipla, utilizando para tal os resultados (2.3.17) e (2.3.18) da mesma subsecção. Deste modo temos no extremo esquerdo[direito] dum intervalo de dominância, comparado o correspondente cultivar dominante com aqueles com coeficiente angular ajustado inferior[superior]. Com efeito, restringi-mo-nos assim aos "níveis mais promissores" nos extremos direito e esquerdo do intervalo de dominância. Se se quiserem considerar todos os cultivares, haverá que alterar os valores críticos das diferenças, para atender ao número de cultivares que se estão a comparar.

Capítulo 3

Análise de um Plano de Melhoramento de Trigo Mole

3.1 Considerações Prévias

O objectivo deste capítulo é mostrar a aplicação dos resultados obtidos nos capítulos anteriores a dados reais, concretamente à análise conjunta dos ensaios, que integram um Plano de Melhoramento de Trigo Mole. As observações de que dispomos encontram-se em suporte informático no apêndice A desta tese e correspondem a ensaios que foram realizados de 1986 a 2000 pela Estação Nacional de Melhoramento de Plantas. Estaremos neste caso, a considerar períodos de tempo mais dilatados, do que aqueles que em geral se consideram, nas aplicações da ACR. Além disso, ao longo do plano, verificou-se a entrada e saída de cultivares. Na Tabela 3.1 encontra-se a lista dos cultivares utilizados, indicando-se, na Tabela 3.2, os anos em que os vários locais foram utilizados. Na Tabela 3.3 tem-se uma linha para cada cultivar, havendo uma coluna para cada ano. A presença [ausência] dum cultivar, num dado ano, é indicada por 1[0].

Tabela 3.1: Lista dos cultivares

anza	sunbird-s	almansor	te9001	te9111	te9406
lima1	neelkant-s	liz1	te9002	te9112	te9503
te8401	miwivet-s	liz2	milan	te9113	te9504
flycatcher	te8601	alva	te9003	te9114	te9712
te8501	te8602	te8801	te8906	te9203	te9713
te8502	te8603	te8802	mondego	te9301	te9714
te8504	te8701	te8901	te9101	te9302	te9715
hahn-s	te8702	te8902	te9102	te9303	te9716

Tabela 3.2: Locais e anos onde se realizaram as experiências

Experiência	Locais	Anos
1	Almeirim	1986/87
2	Évora	1986/88/89
3	Coruche	1986/87
4	Mirandela	1986/89/92
5	Comenda	1986/95/97/99
6	Fundão	1987/88/89
7	E.N.M.P.	1988/89/90/91/92
8	Lamaçais	1988
9	Beja	1988/89/90/91/95/97
10	Benavila	1990/92
11	Elvas	1990
12	Revilheira	1990/95/97/99
13	Santarém	1991
14	Abrantes	1991
15	V.F.Xira	1992/99
16	M.Alhos	2000

Tabela 3.3: Presença e ausência dos cultivares nos vários anos

	1986	1987	1988	1989	1990	1991	1992	1995	1997	1999	2000
anza	1	1	1	1	1	1	1	1	1	1	1
lima1	1	1	1	1	1	1	0	0	0	0	0
te8401	1	0	0	0	0	0	0	0	0	0	0
flycatcher	1	1	1	0	0	0	0	0	0	0	0
te8501	1	1	1	0	0	0	0	0	0	0	0
te8502	1	1	1	0	0	0	0	0	0	0	0
te8504	1	1	1	0	0	0	0	0	0	0	0
hahn-s	1	1	0	0	0	0	0	0	0	0	0
sunbird-s	1	0	0	0	0	0	0	0	0	0	0
neelkant-s	1	0	0	0	0	0	0	0	0	0	0
miwivet-s	1	1	0	0	0	0	0	0	0	0	0
te8601	0	1	1	0	0	0	0	0	0	0	0
te8602	0	1	1	0	0	0	0	0	0	0	0
te8603	0	1	1	1	1	0	0	0	0	0	0
te8701	0	0	1	1	1	0	0	0	0	0	0
te8702	0	0	1	1	1	0	0	0	0	0	0
almansor	0	0	0	1	1	1	1	1	1	1	1
liz1	0	0	0	1	0	0	0	0	0	0	0
liz2	0	0	0	1	1	0	0	0	0	0	0
alva	0	0	0	1	0	0	0	0	0	0	0
te8801	0	0	0	1	1	0	0	0	0	0	0
te8802	0	0	0	1	1	1	0	0	0	0	0
te8901	0	0	0	0	1	1	0	0	0	0	0
te8902	0	0	0	0	1	1	0	0	0	0	0
te9001	0	0	0	0	0	1	0	0	0	0	0
te9002	0	0	0	0	0	1	1	0	0	0	0
milan	0	0	0	0	0	1	0	0	0	0	0
te9003	0	0	0	0	0	1	0	0	0	0	0
te8906	0	0	0	0	0	1	0	0	0	0	0
mondego	0	0	0	0	0	0	1	0	0	0	0
te9101	0	0	0	0	0	0	1	0	0	0	0
te9102	0	0	0	0	0	0	1	0	0	0	0
te9111	0	0	0	0	0	0	1	1	0	0	0
te9112	0	0	0	0	0	0	1	1	0	0	0
te9113	0	0	0	0	0	0	1	1	1	0	0
te9114	0	0	0	0	0	0	1	1	1	0	0
te9203	0	0	0	0	0	0	0	1	1	1	1
te9301	0	0	0	0	0	0	0	1	1	0	0
te9302	0	0	0	0	0	0	0	1	1	0	0
te9303	0	0	0	0	0	0	0	1	1	0	0
te9406	0	0	0	0	0	0	0	1	1	1	1
te9503	0	0	0	0	0	0	0	0	1	1	1
te9504	0	0	0	0	0	0	0	0	1	1	1
te9712	0	0	0	0	0	0	0	0	0	1	1
te9713	0	0	0	0	0	0	0	0	0	1	1
te9714	0	0	0	0	0	0	0	0	0	1	1
te9715	0	0	0	0	0	0	0	0	0	1	1
te9716	0	0	0	0	0	0	0	0	0	1	1

Vamos seguidamente explicar como será conduzida a análise deste plano de melhoramento. Assim numa primeira fase, iremos validar a utilização da ACR no âmbito desta aplicação,

o que faremos atendendo à verificação de duas condições:

1. existência de correlação linear positiva elevada entre os índices ambientais ajustados nas opções *I* e *II*;
2. existência de estabilidade dos cultivares ao longo do plano de melhoramento.

Relativamente à primeira condição, o coeficiente de correlação usual pode ser utilizado como medida da concordância entre os dois conjuntos de estimadores. Se se quiser testar a existência de concordância significativa, pode-se utilizar o coeficiente de correlação ordinal de Spearman, ver Siegel (1988 pág. 235 - 244).

Relativamente à estabilidade dos cultivares, dadas as grandes variâncias dos estimadores das ordenadas na origem, convirá confinar a nossa análise, apenas ao estudo da estabilidade dos coeficientes angulares, ajustados ao longo dos anos, para cada cultivar. Este estudo incide obviamente sobre os cultivares utilizados em mais dum ano. Concluir-se-á pela estabilidade, dos cultivares para os quais não se detectem diferenças significativas. Esses serão designados por β -estáveis.

Para se poder aceitar a utilização da ACR, deverá ter-se alta concordância entre os dois conjuntos de estimadores para os índices ambientais e ainda estabilidade para uma fracção considerável dos cultivares utilizados em mais dum ano. Como veremos adiante, estas condições verificam-se para o plano que estamos a analisar.

Prosseguindo a nossa análise, passaremos à classificação e selecção dos cultivares na Opção *I*. Posteriormente, passaremos à classificação e selecção dos cultivares, utilizando agora os resultados da Opção *II*. Para quaisquer das duas opções os cultivares dividir-se-ão em dominados e dominantes e estes últimos, consoante a posição no seu intervalo de dominância, serão designados por dominantes à direita, intermédios ou dominantes à esquerda. Deste estudo, resultará conhecer-se a evolução do conjunto dos contornos dominantes global e ano a ano, podendo-se estabelecer algumas comparações. Na opção *II* é possível identificar de forma muito clara, certas situações menos objectivas respeitantes à entrada e saída de cultivares do plano. Entre estas, consideramos que a eliminação de cultivares dominantes, para índices ambientais mais baixos, é perfeitamente aceitável. O mesmo não se poderá dizer, quanto à eliminação de cultivares num determinado ano em que os mesmos tinham sido β -estáveis, dominantes à direita e classificados como intensivos. A prática usual consiste em obter cultivares que tenham capacidade de aproveitar índices ambientais elevados, razão pela qual, os cultivares dominantes à direita, devem ser mantidos no programa pela sua contribuição positiva para o alcance de bons rendimentos.

Seguidamente debruça-mo-nos sobre a escolha das funções lineares com que vamos fazer inferência. Trabalharemos com funções da forma $\tilde{U}_j = \tilde{\alpha}_j + \tilde{\beta}_j x$, onde os valores de x a

utilizar pertencem ao intervalo de variação estimado para os índices ambientais, devendo de preferência ser apreciados em conjunto com os melhoradores. É ainda conveniente que os valores de x satisfaçam em simultâneo a condição de invariância de escala, o que explicaremos com mais pormenor adiante. Neste ponto, apresentamos uma primeira contribuição para a fixação desses valores. Vamos ainda identificar o(s) cultivar(es), que em cada ano, foram responsáveis pelos valores máximos tomados pelas funções objectivo lineares escolhidas.

Finalmente, um aspecto que interessa considerar, diz respeito aos cultivares que entraram e saíram do plano de melhoramento nos mesmos anos. Trataremos esses conjuntos de cultivares como grupos. Os cultivares de cada um desses grupos tiveram a mesma "história" dentro do plano. Assim, determinamos que uma correcta condução do plano de melhoramento, exprimir-se-ia pela existência de homogeneidade em cada um desses grupos. Para testar a existência dessa homogeneidade, realizaremos testes para as hipóteses de ausência de efeitos de cultivares e anos. Testaremos ainda a presença de interacções associadas aos pares de níveis destes dois factores. Note-se que, a detecção de faltas de homogeneidade nos grupos de cultivares, poderá por em causa a condução objectiva do plano de melhoramento.

3.2 Índices Ambientais

Utilizámos índices ambientais L_2 para realizar a análise conjunta dos ensaios que integram o Programa de Melhoramento de Trigo Mole. Na Opção *I*, após a realização de seis iterações do algoritmo Zig-Zag, obtivemos os índices ambientais L_2 globais. Na Opção *II*, obtivemos os índices L_2 anuais.¹ Realizaram-se onze ajustamentos com o algoritmo Zig-Zag, um para cada ano, tendo-se atingido a convergência do algoritmo com um número de iterações que variou, para os diversos anos, entre dois e quatro. Note-se que, no plano de melhoramento que estamos a analisar, trabalhamos com um conjunto de cultivares, que em cada ano, se mantém para todos os locais utilizados. Devido a isso, temos neste caso uma rede global anualmente completa. Podemos assim obter índices clássicos, através das médias dos blocos para cada ano.

Como os ensaios tinham quatro repetições, os índices ambientais apresentam-se em grupos de quatro, para qualquer das opções. Na Tabela 3.4 apresentamos as médias dos índices ambientais para esses grupos em Kg por parcela, bem como os índices ambientais estimados nas Opções *I* e *II*, estando os ensaios identificados pelo par (ano, local).

¹Standardizaram-se em cada ano os índices ambientais de modo a que, em todos os anos, o intervalo de variação dos índices ambientais coincida com o mesmo intervalo correspondente ao estudo global (Opção *I*).

Tabela 3.4: Médias dos índices ambientais por bloco

Ensaio	Opção I	Opção II	Índices clássicos
1986 , Almeirim	1.898	1.853	1.860
1986 , Évora	0.759	0.699	0.708
1986 , Coruche	5.486	5.477	5.462
1986 , Mirandela	6.244	6.234	6.222
1986 , Comenda	4.216	4.183	4.148
1987 , Almeirim	4.719	4.843	4.826
1987 , Coruche	4.080	4.248	4.185
1987 , Fundão	4.080	3.970	4.190
1988 , Évora	1.643	1.584	1.592
1988 , Fundão	4.496	4.441	4.463
1988 , ENMP	3.425	3.367	3.380
1988 , Lamaçais	2.770	2.720	2.714
1988 , Beja	3.510	3.448	3.464
1989 , Évora	3.035	2.920	2.948
1989 , Mirandela	6,225	6,025	6.037
1989 , Fundão	2.592	2.491	2.514
1989 , E.N.M.P.	4.328	4.154	4.210
1989 , Beja	4.474	4.30	4.362
1990 , E.N.M.P.	2.402	2.386	2.366
1990 , Beja	2.588	2.558	2.550
1990 , Benavila	0.955	0.945	0.947
1990 , Elvas	0.311	0.311	0.311
1990 Revilheira	0.390	0.351	0.358
1991 , E.N.M.P.	4.746	4.680	4.674
1991 , Beja	3.597	3.509	3.512
1991 , Santarém	1.951	1.837	1.838
1991 , Abrantes	2.261	2.147	2.158
1992 , Mirandela	3.322	3.271	3.311
1992 , E.N.M.P.	1.426	1.783	1.770
1992 , Benavila	2.463	2.410	2.401
1992 , V.F.Xira	4.887	4.816	4.789
1995 , Comenda	1.834	1.859	1.848
1995 , Beja	2.086	2.112	2.115
1995 , Revilheira	1.698	1.665	1.661
1997 , Comenda	3.289	3.314	3.306
1997 , Beja	2.801	2.823	2.839
1997 , Revilheira	1.206	1.16	1.166
1999 , Comenda	3.74	3.828	3.816
1999 , Revilheira	3.00	3.054	3.066
1999 , V.F.Xira	5.134	5.271	5.285
2000 , M.Alhos	1.228	1.542	1.204

Como se observa na Tabela 3.4, os índices ambientais estimados na Opção I, na Opção II e os índices clássicos são muito semelhantes. Os coeficientes de correlação usuais e de Spearman, obtidos para estes pares de índices apresentam-se na Tabela 3.5. Os valores destas correlações estão extremamente próximos de 1, quer sendo obtidas pelo coeficiente de correlação linear, quer pelo coeficiente de correlação ordinal de Spearman. Verifica-se pois, a presença de fortes relações lineares positivas entre os três tipos de índices. Os valores

Tabela 3.5: Correlações lineares e de Spearman entre pares de índices

Par	Coefficiente de correlação linear	Coefficiente de correlação de Spearman
L_2 global, L_2 anual	0.9948	0.9648
L_2 global, Índices clássicos	0.9957	0.9818
L_2 anual, Índices clássicos	0.9988	0.9820

apresentados para os coeficientes de correlação de Spearman são, ver Siegel (1988, pág. 235 - 244) significativos ao nível de 1%. Previam-se estes resultados, ver Pereira (2004), entre as séries de valores estimados para os índices L_2 anuais e clássicos. Relativamente ao par (L_2 global, L_2 anual) é de realçar o registo de correlações muito acentuadas, o que de certo modo valida a utilização da ACR no âmbito desta aplicação.

Veja-se ainda que, se se tiver

$$Z_{j+1} \approx \gamma_j + \eta_j Z_j, j = 1, 2 \quad (3.2.1)$$

ter-se-á

$$Z_3 \approx \gamma_2 + \eta_2(\gamma_1 + \eta_1 Z_1) \approx (\gamma_2 + \eta_2 \gamma_1) + \eta_2 \eta_1 Z_1 \quad (3.2.2)$$

pelo que a existência das duas relações lineares referidas, implicam a existência duma terceira relação linear, entre L_2 global e índices clássicos.

3.3 Estudo da estabilidade dos cultivares

Ainda na fase de validar a utilização da ACR no Plano de Melhoramento do Trigo, vamos verificar se os cultivares utilizados apresentam estabilidade. Começamos por apresentar na Tabela 3.6, para os vários cultivares, os coeficientes angulares das rectas de regressão ajustadas nas opções *I* e *II*. Para a Opção *II* indica-se o ano.

Tabela 3.6: Coeficientes angulares ajustados nas Opções *I* e *II*

Cultivares	L_2 global	1986	1987	1988	1989	1990	1991	1992	1995	1997	1999	2000
Anza	1.051	1.097	0.258	0.994	0.989	0.781	0.992	1.070	1.289	1.059	0.872	0.893
Lima1	1.082	1.012	0.79	1.268	1.293	1.017	1.082					
Te8401	1.014	1.005										
Flycatcher	1.037	1.025	0.616	1.08								
Te8501	1.092	1.151	1.038	0.905								
Te8502	1.089	1.029	2.411	1.232								
Te8504	0.92	0.937	0.032	0.841								
Hahn-s	1.035	1.015	2.216									
Sunbird-s	0.811	0.906										
Neelkant-s	0.862	0.855										
Miwivet-s	1.048	1.010	1.105									
Te8601	1.02		0.825	1.064								
Te8602	1.123		0.881	1.080								
Te8603	0.896		0.721	0.851	0.847	1.04						
Te8701	0.836			0.751	0.885	0.925						
Te8702	0.971			1.010	1.053	0.905						
Almansor	0.987				1.076	0.949	0.997	1.013	0.744	0.918	0.843	1.46
Liz1	0.93				0.955							
Liz2	1.002				0.923	0.966						
Alva	1.048				1.074							
Te8801	0.925				0.936	1.032						
Te8802	0.961				0.948	0.987	0.939					
Te8901	1.033					1.112	1.10					
Te8902	1.034					1.205	0.933					
Te9001	0.999						0.98					
Te9002	0.994						0.981	0.997				
Milan	0.869						0.853					
Te9003	1.18						1.128					
Te8906	0.967						0.949					
Mondego	1.186							1.195				
Te9101	0.645							0.651				
Te9102	0.918							0.925				
Te9111	0.91							0.819	1.498			
Te9112	0.983							0.973	0.759			
Te9113	1.056							1.095	0.789	0.937		
Te9114	1.144							1.205	0.863	0.896		
Te9203	1.01								0.494	1.131	0.936	2.926
Te9301	1.103								1.1898	1.087		
Te9302	0.628								1.586	0.774		
Te9303	1.039								1.535	0.993		
Te9406	1.245								0.7533	1.142	1.44	0.939
Te9503	0.939									1.056	1.019	0.994
Te9504	1.037									0.978	1.235	0.927
Te9712	1.087										0.961	1.458
Te9713	1.116										1.209	1.107
Te9714	1.071										0.785	1.107
Te9715	1.046										0.876	-5.166
Te9716	0.933										0.819	-0.864

Observe-se que, a haver proximidade entre os coeficientes ajustados pelas duas Opções, os coeficientes ajustados ano a ano tenderão a oscilar em torno dos obtidos pelo ajustamento global. Como podemos verificar por (2.3.3) da subsecção 2.3.2, quando $\tilde{x}_{.j}$ não é pequeno,

$\tilde{Var}(\tilde{\alpha}_j), j = 1, \dots, J$ é muito maior que $\tilde{Var}(\tilde{\beta}_j), j = 1, \dots, J$, o que nos indica que deveremos apenas considerar a oscilação dos coeficientes angulares. No nosso estudo, a variável controlada é o índice ambiental e verifica-se que os valores das somas dos quadrados dos desvios para a média dos índices ambientais, s , variam muito de ano para ano, tal como revelam os resultados apresentados na Tabela 3.7.

Tabela 3.7: Soma dos quadrados dos Resíduos para a média dos índices ambientais obtidos na Opção II

Ano	1986	1987	1988	1989	1990	1991	1992	1995	1997	1999	2000
s	90.588	2.902	19.198	33.33	19.639	22.237	21.949	0.614	10.6	10.46	0.064

Decidiu-se assim pela exclusão dos anos de 1987, 1995 e 2000 por falta de variação dos índices L_2 ajustados nos mesmos anos. Ficamos reduzidos a 22 cultivares, utilizados em pelo menos dois dos anos não excluídos. Esses cultivares são: Anza, Lima1, Flycatcher, Te8501, Te8502, Te8504, Te8603, Te8701, Te8702, Almansor, Liz2, Te8801, Te8802, Te8901, Te8902, Te9002, Te9113, Te9114, Te9203, Te9406, Te9503, Te9504. Vamos testar, cultivar a cultivar, a hipótese de igualdade dos coeficientes angulares para os anos em que foram utilizados. Para tal, utilizamos os resultados dados pela subsecção 2.3.3, nomeadamente a Proposição 2.8. Apresentam-se na Tabela 3.8 os resultados dos testes F .

Tabela 3.8: Resultados dos testes F

Cultivar	Valor do teste F
Anza	1.504
Lima1	3.418 (*)
Flycatcher	0.246
Te8501	5.000 (*)
Te8502	3.387
Te8504	0.762
Te8603	1.356
Te8701	0.862
Te8702	0.709
Almansor	0.495
Liz2	0.116
Te8801	0.590
Te8802	0.070
Te8901	0.012
Te8902	4.020 (*)
Te9002	0.015
Te9113	0.928
Te9114	3.550
Te9203	1.044
Te9406	2.429
Te9503	0.037
Te9504	1.807

Os valores significativos ao nível de 5% encontram-se assinalados com o símbolo (*). Dado haver apenas três casos com diferenças significativas, em 22 cultivares considerados, estamos perante um bom resultado. A utilização da ACR está deste modo validada, podendo-se assim dar seguimento à análise.

3.4 Classificação dos cultivares na Opção I

Vamos começar por classificar os cultivares no âmbito da Opção I. Utilizaremos para testar as hipóteses os valores obtidos no ajustamento global. Como foi visto na subsecção 2.4.1, quando se tem $\tilde{\beta}_j > 1$, testa-se

$$H_{o,\beta_j} : \beta_j = 1 \quad (3.4.1)$$

contra

$$H_{1,\beta_j} : \beta_j > 1 \quad (3.4.2)$$

e, se $\tilde{\beta}_j < 1$, testa-se

$$H_{o,\beta_j} : \beta_j = 1 \quad (3.4.3)$$

contra

$$H_{2,\beta_j} : \beta_j < 1 \quad (3.4.4)$$

Apresentamos na Tabela 3.9 os valores estimados para $\tilde{\alpha}_j$, $\tilde{\beta}_j$, $\tilde{\alpha}'_j = \tilde{\alpha}_j + \tilde{\beta}_j\theta_o$, $j = 1, \dots, 48$ ordenados pela ordem decrescente dos valores de $\tilde{\alpha}'_j$ e na Tabela 3.10 os resultados dos testes relativos às hipóteses acima enunciadas. O ponto médio do intervalo de variação estimado para os índices ambientais é de $\theta_o = 3.416$.

Tabela 3.9: Valores estimados para α , β e α'

Cultivar	$\bar{\alpha}$	$\tilde{\beta}$	$\bar{\alpha}'$
te9406	-0.393	1.245	3.862
te9713	0.006	1.116	3.818
te9114	-0.151	1.144	3.756
miwivet-s	0.155	1.048	3.736
te8602	-0.182	1.123	3.652
hahn-s	0.117	1.035	3.652
te8902	0.074	1.034	3.605
te9714	-0.061	1.071	3.599
te8901	0.056	1.033	3.584
te9102	0.448	0.918	3.582
te9301	-0.192	1.103	3.577
te9203	0.116	1.010	3.566
te9712	-0.260	1.087	3.555
te9716	0.354	0.933	3.539
te8502	-0.189	1.089	3.532
lima1	-0.177	1.082	3.520
te9303	-0.047	1.039	3.504
mondego	-0.562	1.186	3.491
te9003	-0.543	1.180	3.489
te9113	-0.135	1.056	3.472
liz1	0.293	0.930	3.471
anza	-0.143	1.051	3.445
liz2	0.011	1.002	3.435
te9002	0.031	0.994	3.428
te8501	-0.304	1.092	3.427
te8401	-0.052	1.014	3.411
te8601	-0.080	1.020	3.404
te8603	0.327	0.896	3.387
almansor	-0.014	0.987	3.356
te9715	-0.227	1.039	3.345
alva	-0.238	1.048	3.340
flycatcher	-0.214	1.037	3.330
te8802	0.044	0.961	3.328
te9504	-0.228	1.037	3.316
te9112	-0.042	0.983	3.315
te8504	0.164	0.920	3.307
te9001	-0.140	0.999	3.272
te8801	0.061	0.925	3.221
te9503	-0.044	0.939	3.164
te8702	-0.199	0.971	3.119
neelkant-s	0.172	0.869	3.117
te9111	-0.020	0.910	3.090
te9302	0.817	0.628	2.962
te8701	0.1	0.836	2.956
milan	-0.026	0.869	2.944
te9101	0.486	0.645	2.690
sunbird-s	-0.105	0.811	2.665
te8906	-0.655	0.967	2.649

Tabela 3.10: Classificação dos cultivares na Opção I

$\beta < 1$	$\beta < 1$	$\beta > 1$	$\beta > 1$	classificação
significativo	não significativo	não significativo	significativo	
			te9406	intensivo
		te9713		estável
		te9114		estável
		miwivet		estável
		te8602		estável
		hahn-s		estável
		te8902		estável
		te9714		estável
		te8901		estável
	te9102			estável
		te9301		estável
		te9203		estável
	te9716			estável
			te8502	intensivo
			lima 1	intensivo
		te9303		estável
			mondego	intensivo
			te9003	intensivo
		te9113		estável
	liz1			estável
		te9712		estável
			anza	intensivo
		liz2		estável
	te9002			estável
			te8501	intensivo
		te8401		estável
		te8601		estável
te8603				extensivo
	almansor			estável
		te9715		estável
		alva		estável
		flycatcher		estável
	te8802			estável
		te9504		estável
	te9112			estável
te8504				extensivo
	te9001			estável
te8801				extensivo
	te9503			estável
	te8702			estável
neelkant-s				extensivo
	te9111			estável
te9302				extensivo
te8701				extensivo
	milan			estável
te9101				extensivo
sunbird-s				extensivo
	te8906			estável

No parágrafo seguinte apresentamos os resultados do contorno superior na Opção *I*, após o que faremos alguns comentários relativos aos resultados da classificação e da selecção na Opção *I*, em simultâneo.

3.5 Análise do Contorno Superior na opção *I*

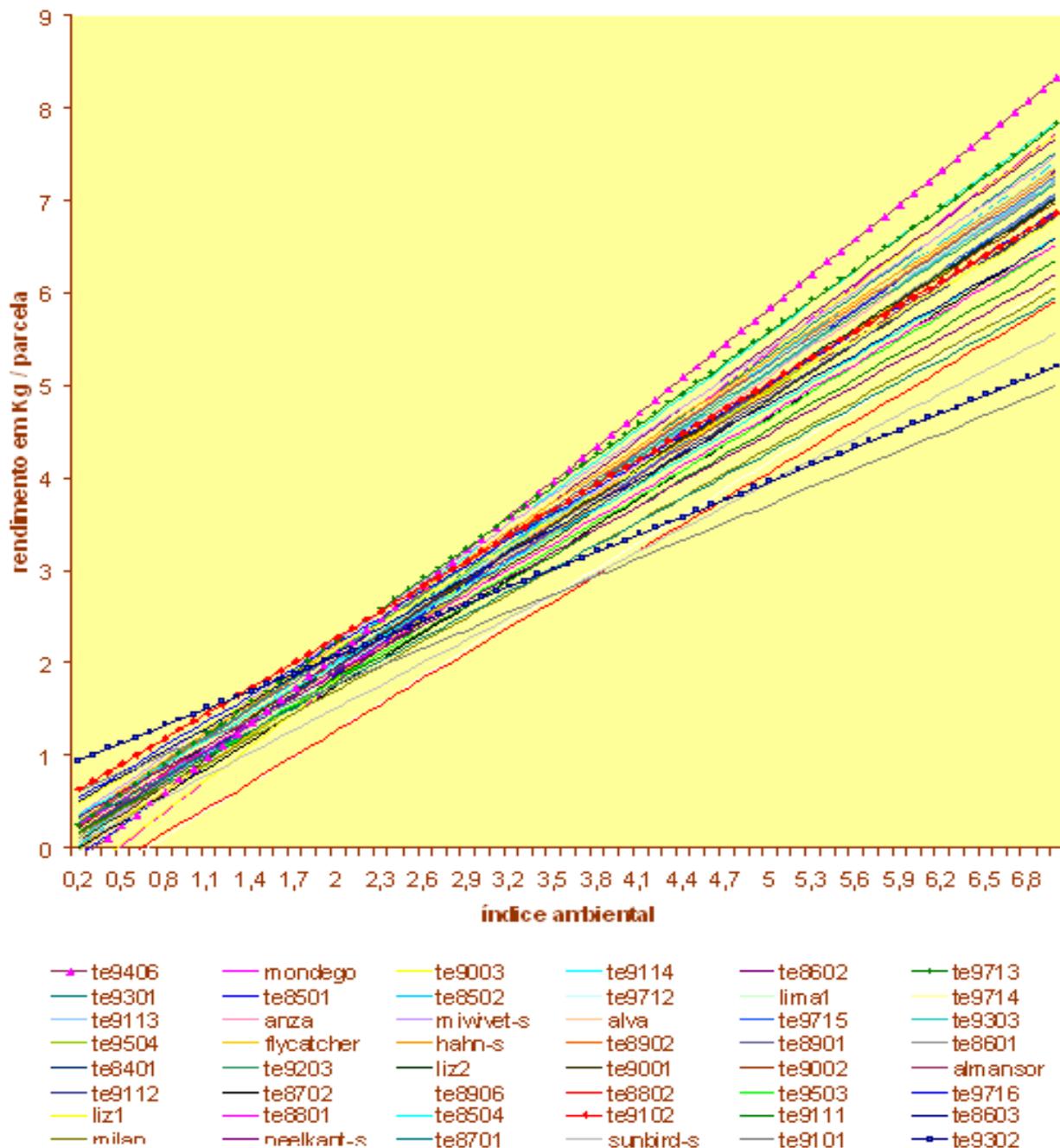
Na Tabela 3.11 apresentam-se os coeficientes ajustados e os valores de R^2 para as regressões ajustadas na Opção *I*.

Tabela 3.11: Resultados do ajustamento

Cultivar	$\hat{\alpha}$	$\hat{\beta}$	R^2
te9406	-0.393	1.245	0.94
Mondego	-0.562	1.186	0.87
te9003	-0.543	1.180	0.97
te9114	-0.151	1.144	0.93
te8602	-0.182	1.123	0.86
te9713	0.006	1.116	0.96
te9301	-0.192	1.103	0.94
te8501	-0.304	1.092	0.92
te8502	-0.189	1.089	0.88
te9712	-0.260	1.087	0.96
limal	-0.177	1.082	0.92
te9714	-0.061	1.071	0.92
te9113	-0.135	1.056	0.94
anza	-0.143	1.051	0.92
miwivet-s	0.155	1.048	0.89
alva	-0.238	1.048	0.92
te9715	-0.227	1.039	0.94
te9303	-0.047	1.039	0.95
te9504	-0.228	1.037	0.87
flycatcher	-0.214	1.037	0.94
hahn-s	0.117	1.035	0.85
te8902	0.074	1.034	0.94
te8901	0.056	1.033	0.94
te8601	-0.080	1.020	0.90
te8401	-0.052	1.014	0.97
te9203	0.116	1.010	0.92
liz2	0.011	1.002	0.96
te9001	-0.140	0.999	0.94
te9002	0.031	0.994	0.89
almansor	-0.014	0.987	0.94
te9112	-0.042	0.983	0.96
te8702	-0.199	0.971	0.95
te8906	-0.655	0.967	0.97
te8802	0.044	0.961	0.97
te9503	-0.044	0.939	0.87
te9716	0.354	0.933	0.91
liz1	0.293	0.930	0.91
te8801	0.061	0.925	0.95
te8504	0.164	0.920	0.86
te9102	0.448	0.918	0.93
te9111	-0.020	0.910	0.85
te8603	0.327	0.896	0.89
milan	-0.026	0.869	0.88
neelkant-s	0.172	0.869	0.9
te8701	0.1	0.836	0.91
sunbird-s	-0.105	0.811	0.90
te9101	0.486	0.645	0.56
te9302	0.817	0.628	0.46

Observe-se que apenas os dois últimos cultivares apresentam valores mais baixos para R^2 , podendo considerar-se que na generalidade se ajustaram bem as regressões. Na figura 3.1 apresentamos as rectas de regressão ajustadas para os diferentes cultivares, identificando-se facilmente os cultivares que compõem o contorno superior.

Figura 3.1: Contorno Superior na Opção I



Da análise da figura 3.1 resultam os seguintes cultivares dominantes:

Cultivar dominante	região de dominância	Anos de utilização do cultivar
te9406	[3.08 ; 6.52]	1995,1997,1999,2000
te9713	[2.23 ; 3.08[1999,2000
te9102	[1.28 ; 2.23[1992
te9302	[0.31 ; 1.28[1995,1997

Verificamos que, os quatro cultivares dominantes, foram introduzidos no processo de selecção após 1992. Destes, dois são dominantes para índices ambientais baixos, te9102 e te9302, sendo rapidamente eliminados do processo de selecção. Os outros dois cultivares dominantes, te9406 e te9713, foram introduzidos em 1995 e 1999 respectivamente, sendo mantidos até ao final. É interessante ver-se que, os cultivares dominantes para índices ambientais mais elevados, foram mantidos até ao final do programa. Salientamos ainda como importante o facto do cultivar te9406, β -estável, dominante à direita, ter sido classificado como intensivo. Este facto revela que existe coerência entre as técnicas de classificação e selecção de cultivares. É importante observar que os cultivares dominantes à direita tendem a ser, como neste caso, significativamente intensivos. Os cultivares te9713 e te9102 são classificados como estáveis, sendo β -estáveis e dominantes intermédios. Já o cultivar dominante para ambientes mais "pobres", te9302, é classificado como extensivo.

Vamos agora proceder às comparações, dos cultivares dominantes com os não dominantes, nos pontos de separação das respectivas regiões de dominância. Assim,

1. no ponto $\theta = 3.08$, compara-se o cultivar te9406, com todos aqueles para os quais o declive da recta de regressão ajustada é inferior. Neste mesmo ponto, deve comparar-se cultivar te9713, com todos aqueles cuja recta ajustada apresenta um declive superior;
2. no ponto $\theta = 2.23$, compara-se o cultivar te9713, com todos aqueles para os quais a recta ajustada tem declive inferior, e compara-se o cultivar te9102, com todos aqueles cuja recta ajustada tem declive superior;
3. no ponto $\theta = 1.26$, compara-se o cultivar te9102, com todos aqueles para os quais a recta ajustada apresenta declive inferior e compara-se o cultivar te9302, com todos aqueles cuja recta ajustada apresenta declive superior.

Na Tabela 3.12 apresentamos os resultados obtidos pela realização de testes t unilaterais ao nível de significância de 5% para as hipóteses,

1. $H_{o,1,j'}(3.08) : U_1(3.08) = U_{j'}(3.08), j' = 2, \dots, 48;$
2. $H_{o,6,j'}(3.08) : U_6(3.08) = U_{j'}(3.08), j' = 1, \dots, 5;$
3. $H_{o,6,j'}(2.23) : U_6(2.23) = U_{j'}(2.23), j' = 7, \dots, 48;$
4. $H_{o,40,j'}(2.23) : U_{40}(2.23) = U_{j'}(2.23), j' = 41, \dots, 48;$
5. $H_{o,40,j'}(1.26) : U_{40}(1.26) = U_{j'}(1.26), j' = 1, \dots, 39;$
6. $H_{o,48,j'}(1.26) : U_{48}(1.26) = U_{j'}(1.26), j' = 1, \dots, 47.$

tal como se indicou na subsecção 2.3.3.

Tabela 3.12: Resultados do teste t - Cultivares dominantes e dominados

Cultivar dominante	região de dominância	cultivares dominados ao nível de 5%
te9406	[3.08 ; 6.524]	mondego, te9003, te8501, te8502 , te9712, lima1, te9113, anza alva, te9715, te9303, te9504 flycatcher, te8901, te8601, te8401 te9203, liz2, te9001,te9002 almansor, te9112,te8702, te8906 te8802, te9503, liz1, te8801 te8504, te9111, te8603, milan , neelkant-s,te8701
te9713	[2.23 ; 3.08]	mondego, te9003, te8501, te9712 lima1, te9113, anza, alva te9715, te9504, flycatcher, te9001 almansor, te9112, te8702, te8906 te8802 te9503, te8801, te8504 te8701, sunbird-s, te9101, te9302 te9111, milan, neelkant-s
te9102	[1.27 ; 2.23]	mondego, te9003, te8501, te9712 lima1, te9113, anza, alva te9715, te9504, flycatcher, almansor te9112, te8702, te8906, te8802 te9503, te8801, te8504, te9111 milan, neelkant-s, te8701, sunbird-s
te9302	[0.31 ; 1.27]	te9406, mondego, te9003, te9114 te9301, te8501, te8502, te9712 lima1, te9113, anza, alva te9715, te9504, flycatcher, te8401 liz2, te9001, almansor, te9112 te8702, te8906, te8802, te9503 te9111, milan, neelkant-s, te8701 sunbird-s

Apresentamos na Tabela 3.13 os resultados dos métodos de comparação múltipla de Bonferroni e de Scheffé também ao nível de significância de 5%.

Tabela 3.13: Resultados dos métodos de comparação múltipla de Bonferroni e de Scheffé ao nível de 5%

Método	te9302	te9102	te9102	te9713	te9713	te9406
	$\theta_{o,1} = 1.275$	$\theta_{o,1} = 1.275$	$\theta_{o,2} = 2.227$	$\theta_{o,2} = 2.227$	$\theta_{o,3} = 3.08$	$\theta_{o,3} = 3.08$
Scheffé	a)	a)	a)	a)	a)	te8906 sunbird-s
Bonferroni	te8702 te8906	te9111 te8701 sunbird-s	te8702 te8906	te8702 te8701 te8906 sunbird-s	a)	te8501 anza te9715 te9504 flycatcher almansor te9112 te8702 te8906 te8802 te9503 te8801 te8504 te9111 te8603 milan neelkant-s te8701 sunbird-s te9101 te9302

Por exemplo, utilizando o método de Bonferroni, conclui-se, ao nível de 5 %, que o cultivar te9302 domina, no respectivo intervalo, os cultivares te8702 e te8906. O símbolo *a)* indica ausência de diferenças significativas.

Relativamente às técnicas estatísticas utilizadas nestas comparações, podemos concluir que:

1. o método de Scheffé é muito conservador;
2. o método de Bonferroni é uma solução intermédia entre o teste *t* e o método de Scheffé, pelo que se recomenda a sua utilização.

3.6 Classificação dos cultivares na Opção II

Para facilitar a interpretação dos resultados, vamos para a Opção II, apresentar a classificação dos cultivares de um modo diferente ao usado na Opção I. Assim, na Tabela 3.14, os números 1, 2 e 3 significam intensivo, estável e extensivo respectivamente .

Tabela 3.14: Classificação dos cultivares na Opção II

	1986	1987	1988	1989	1990	1991	1992	1995	1997	1999
Anza	1	3	2	2	3	2	2	2	2	2
Lima1	2	2	1	1	2	2				
Te8401	2									
neelkant	3									
sunbird	3									
Te8501	1	2	2							
Te8502	2	1	1							
Flycatcher	2	2	2							
Te8504	2	3	2							
hanh-s	2	1								
miwivet	2	2								
Te8601		2	2							
Te8602		2	2							
Te8603		2	2	3	2					
Te8701			3	2	2					
Te8702			2	2	2					
Almansor				2	2	2	2	2	2	2
Liz1				2						
Liz2				2	2					
Alva				2						
Te8801				2	2					
Te8802				2	2	2				
Te8901					2	2				
Te8902					1	2				
Milan						2				
Te9003						1				
Te9001						2				
Te8906						2				
Te9002						2	2			
Mondego							1			
Te9101							3			
Te9102							2			
Te9111							3	2		
Te9112							2	2		
Te9113							2	2	2	
Te9114							1	2	2	
Te9203								2	2	2
Te9406								2	2	1
Te9301								2	2	
Te9302								2	2	
Te9303								2	2	
Te9503									2	2
Te9504									2	2
Te9712										2
Te9713										2
Te9714										2
Te9715										2
Te9716										2

Ao observar a tabela 3.14 é possível verificar as oscilações da classificação, para cada cultivar, ao longo dos anos. Faremos alguns comentários mais adiante, após a apresentação dos resultados da selecção na Opção II.

3.7 Análise do contorno superior na Opção II

A análise dos contornos superiores, obtidos através dos ajustamentos ano a ano, permite-nos avaliar como foi sendo conduzido o plano de melhoramento. Dado estarmos a considerar

Após o exame das tabelas 3.14 e 3.15, salientamos que:

1. existem grupos de cultivares com a mesma história dentro do plano, adiante analisaremos esses grupos;
2. três cultivares (Anza, lima1 e Almansor), tiveram uma longa permanência no programa;
3. o cultivar miniwet foi dominante intermédio em 1987, tendo sido eliminado nesse ano;
4. os cultivares te8502 e lima1 foram eliminados respectivamente nos anos de 1988 e 1991, anos em que os mesmos foram dominantes à direita, β -estáveis e classificados como intensivos;
5. O cultivar te9406 é dominante à direita no contorno superior global, é também dominante à direita no ano de 1999, sendo ainda um cultivar que apresenta um valor elevado para α' , classificado como intensivo e β -estável;
6. Note-se que nos anos de 1986, 1987, 1988, 1989, 1990, 1992 e 1999, o cultivar dominante à direita foi sempre classificado como intensivo, e em 1991, 1995 e 1997 como estável. Os anos de 1995 e 1997 são anos com poucos ensaios, o que explica a falta de potência do teste utilizado na classificação dos cultivares na Opção II.

3.8 Selecção das funções objectivo

Como vimos na subsecção 2.3.2, para se estabelecerem certas comparações relativas a diferentes anos, torna-se vantajoso utilizar como funções objectivo, funções lineares invariantes,

$$\tilde{U}_j(1, x) = \tilde{\alpha}_j + \tilde{\beta}_j x \quad (3.8.1)$$

com

$$x = r\tilde{x}^+ + (1 - r)\tilde{x}^-, 0 < r < 1 \quad (3.8.2)$$

excluindo-se portanto, a opção aparentemente conveniente de igualar x à média dos índices ambientais ajustados.

Dado pretender-se cultivares capazes de aproveitar boas capacidades produtivas, pode-se começar por tomar $r = 1$.

Por outro lado, se se quiser comparar comportamentos médios num intervalo, $[x_1, x_2]$ com,

$$x_j = r_j x^+ + (1 - r_j) x^-, j = 1, 2 \quad (3.8.3)$$

como

$$\frac{1}{(x_2 - x_1)} \int_{x_1}^{x_2} (\tilde{\alpha}_j + \tilde{\beta}_j x) dx = \tilde{\alpha}_j + \tilde{\beta}_j \frac{x_1 + x_2}{2} \quad (3.8.4)$$

pode utilizar-se

$$x = \frac{x_1 + x_2}{2}. \quad (3.8.5)$$

Consideramos que os valores de r devem ser fixados de acordo com os melhoradores. Como primeira aproximação, propomos os valores que apresentamos na Tabela 3.16, pois para estes, as funções lineares que utilizamos, são invariantes à mudança de escala.

Tabela 3.16: Funções objectivo utilizadas

r	x	$\alpha + \beta x$
1	6.524	$f_1(x) = \alpha + \beta 6.524$
$\frac{2}{3}$	4.452	$f_2(x) = \alpha + \beta 4.452$
$\frac{1}{2}$	3.416	$f_3(x) = \alpha + \beta 3.416$

Estas funções correspondem a valores de x que foram obtidos por uma combinação linear das extremidades do intervalo de variação para os índices ambientais, $x^+ = 6.524$ e $x^- = 0.309$.

Observe-se que sendo,

$$x = rx^+ + (1 - r)x^- \quad (3.8.6)$$

o ponto médio dum intervalo com extremo direito x^+ , para que o extremo esquerdo não seja inferior a x^- , terá de se ter

$$x^+ - (rx^+ + (1 - r)x^-) \leq (rx^+ + (1 - r)x^-) - x^-, \quad (3.8.7)$$

o que é equivalente a ter-se $r \geq \frac{1}{2}$, condição que define os casos que na prática interessa considerar. Aliás este valor limite foi um dos escolhidos.

3.9 Máximos das funções objectivo

Na Tabela 3.17 encontram-se indicados, ano a ano, os valores máximos das funções objectivo utilizadas e o cultivar para os quais os mesmos foram obtidos.

Tabela 3.17: Máximos das funções objectivo

Ano	f_1	f_2	f_3
1986	te8501	hahn	hahn
	7.139	5.036	4.04
1987	te8502	miwivet	miwivet
	7.036	5.23	4.787
1988	te8502	te8603	te8603
	5.204	3.855	3.381
1989	lima1	lima1	anza
	7.264	5.376	4.470
1990	te8902	te8902	te8901
	3.153	2.191	1.723
1991	lima1	lima1	lima1
	5.876	4.466	3.761
1992	te9114	te9114	te9114
	5.796	4.385	3.68
1995	te9302	te9302	te9302
	2.852	2.487	2.305
1997	te9203	te9203	almansor
	3.679	2.786	2.359
1999	te9406	te9406	te9406
	6.492	5.336	4.758

Analisando os resultados do contorno superior na Opção *II*, em simultâneo com os resultados da Tabela 3.17, verifica-se que o plano de melhoramento foi conduzido com algumas deficiências, das quais salientamos:

1. o cultivar Lima 1 foi eliminado em 1991, ano em que foi dominante à direita e responsável pelos valores máximos tomados pelas funções f_1 , f_2 e f_3 ;
2. O cultivar Te8502 foi eliminado após ter sido dominante à direita dois anos consecutivos. Este cultivar foi também responsável pelo valor máximo assumido pela

função f_1 em 1987 e 1988;

3. Verifica-se uma grande variabilidade nos valores máximos das funções objectivo ao longo dos anos. Essa variabilidade estará possivelmente ligada a variações climáticas. Adiante voltaremos a esta questão.

3.10 Análise dos grupos de Cultivares

No plano de melhoramento que temos vindo a analisar, há grupos de cultivares que aparecem sempre nos mesmos pares (local, ano), como se observa na Tabela 3.18.

Tabela 3.18: Agrupamento dos cultivares segundo a presença nos anos

Índices dos Grupos	Composição dos grupos
1	Anza
2	Lima1
3	te8401, sunbird-s, neelkant
4	flycatcher, te8501, te8502, te8504
5	hahn's, miwivet-s
6	te8601, te8602
7	te8603
8	te8701, te8702
9	almansor
10	liz1, alva
11	liz2, te8801
12	te8802
13	te8901, te8902
14	te9001, milan, te9003, te8906
15	te9002
16	mondego, te9101, te9102
17	te9111, te9112
18	te9113, te9114
19	te9203, te9406
20	te9301, te9302, te9303
21	te9503, te9504
22	te9712, te9713, te9714, te9715, te9716

Os cultivares que constituem estes grupos, entraram e saíram do programa de melhoramento nos mesmos anos. Uma correcta condução do plano de melhoramento, deveria ter ditado a saída simultânea de cultivares com o mesmo comportamento no programa. Vamos utilizar os resultados da subsecção 2.3.4, para realizar análises de variância, grupo a grupo, para os valores tomados pelas funções objectivo. Se um grupo foi apenas utilizado num ano, o único factor a considerar terá como níveis os cultivares do grupo. Se o grupo foi utilizado em vários anos, temos para além desse, um factor cujos níveis são os anos e a interacção entre eles.

Remetemos para o apêndice A as tabelas que contêm as observações que foram utilizadas na análise de cada grupo. Vamos exemplificar com a aplicação das estatísticas F dadas por (2.3.40) aos dados do grupo quatro com a função objectivo f_1 . Na Tabela 3.19 apresentam-se os dados sujeitos a comparação para este exemplo. Na sexta coluna da mesma tabela temos os valores estimados para as variâncias. Assim, por exemplo, o valor 0.138, representa a variância estimada para a função $f_1 = \alpha_j + \beta_j x$, com $j = 1, \dots, 4$, no ano de 1986.

Tabela 3.19: Valores estimados para a função f_1 relativos aos cultivares do Grupo quatro

Grupo 4	flycatcher	te8501	te8502	te8504	$v\tilde{a}r$
1986	6.537	7.139	6.719	6.187	0.138
1987	4.927	5.222	7.036	4.303	0.477
1988	4.727	4.593	5.204	4.334	0.174

Para este caso as hipóteses a testar são as de ausência de efeitos

$$\begin{cases} H_{o,\gamma} : \gamma_j = 0; j = 1, 2, 3, 4 \\ H_{o,\eta} : \eta_l = 0; l = 1, 2, 3 \end{cases} \quad (3.10.1)$$

e de ausência de interacções

$$H_{o,\lambda} : \lambda_{j,l} = 0; j = 1, 2, 3, 4; l = 1, 2, 3 \quad (3.10.2)$$

sendo os efeitos dos níveis dos dois factores e as interacções entre pares de níveis dadas por,

$$\begin{cases} \gamma_j = U_{j,\cdot} - U_{\cdot\cdot}, & j = 1, 2, 3, 4 \\ \eta_l = U_{\cdot,l} - U_{\cdot\cdot}, & l = 1, 2, 3 \\ \lambda_{j,h} = U_{j,h} - U_{j,\cdot} - U_{\cdot,h} + U_{\cdot\cdot}, & j = 1, 2, 3, 4; l = 1, 2, 3 \end{cases} \quad (3.10.3)$$

com

$$\begin{cases} U_{j,\cdot} = \frac{1}{3} \sum_{l=1}^3 U_{j,l}, j = 1, 2, 3, 4 \\ U_{\cdot,l} = \frac{1}{4} \sum_{j=1}^4 U_{j,l}, l = 1, 2, 3 \\ U_{\cdot\cdot} = \frac{1}{12} \sum_{j=1}^4 \sum_{l=1}^3 U_{j,l} \end{cases} \quad (3.10.4)$$

e

$$\tilde{f}_1 = \tilde{U}_{j,l}(6.524) = \tilde{\alpha}_{j,l} + 6.524\tilde{\beta}_{j,l}, j = 1, 2, 3, 4; l = 1, 2, 3. \quad (3.10.5)$$

Começamos por obter as matrizes \mathbf{L}_3 , \mathbf{L}_4 , \mathbf{A}_γ , \mathbf{A}_η e \mathbf{A}_λ , que serão por exemplo

$$\mathbf{L}_3 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \quad (3.10.6)$$

$$\mathbf{L}_4 = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{6} & -\frac{\sqrt{3}}{6} & -\frac{\sqrt{3}}{6} \\ 0 & \frac{\sqrt{6}}{3} & -\frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{6} \\ 0 & 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \quad (3.10.7)$$

$$\begin{cases} \mathbf{A}_4 & = \mathbf{L}_4 \otimes \frac{1}{\sqrt{3}} \mathbf{1}'_3 \\ \mathbf{A}_3 & = \frac{1}{\sqrt{4}} \mathbf{1}'_4 \otimes \mathbf{L}_3 \\ \mathbf{A}_{12} & = \mathbf{L}_4 \otimes \mathbf{L}_3 \\ & . \end{cases} \quad (3.10.8)$$

No vector $\tilde{\mathbf{U}}_{12}$, as primeiras três posições são ocupadas pelas observações correspondentes ao cultivar flycatcher, as três seguintes ao cultivar te8501 e assim sucessivamente até ao último cultivar. Ter-se-á ainda a matriz diagonal das variâncias estimadas,

$$\mathbf{D} = \mathit{diag}[0.138, 0.477, 0.174, 0.138, 0.477, 0.174, 0.138, 0.477, 0.174, 0.138, 0.477, 0.174] \quad (3.10.9)$$

e

$$\frac{\tilde{S}}{g} = \frac{293.7252}{1526} = 0.192. \quad (3.10.10)$$

Após alguns cálculos matriciais, obtêm-se os valores para as estatísticas de teste

$$\begin{cases} F_\gamma = 19.897 \\ F_\eta = 128.132 \\ F_\lambda = 5.635. \end{cases} \quad (3.10.11)$$

Qualquer destas estatísticas é significativa ao nível de 5%, pelo que podemos concluir que as observações deste grupo apresentam diferenças significativas devidas aos efeitos dos factores cultivar, ano e às interacções. Os cultivares do grupo quatro, avaliados pela função f_1 , não apresentam assim um comportamento homogéneo.

Apresentamos na Tabela 3.20 os valores obtidos para as estatísticas de teste para todos os grupos de cultivares que sejam constituídos por dois ou mais cultivares. Também nesta análise excluimos o ano 2000 pelo mesmo motivo que já referimos anteriormente. Assinalamos com o símbolo (*) os casos para os quais, as hipóteses nulas foram rejeitadas, ao nível de significância de 5%.

Tabela 3.20: Resultados dos Testes F para efeitos e interações

função	Grupo	Cultivares	Anos	Interação	Grupo	Cultivares	Anos	Interação
f_1		17.825*				5.715*		
f_2	Grupo 3	21.507*			Grupo 14	9.019*		
f_3		16.179*				10.540*		
f_1		19.897*	128.13*	5.635*		17.842*		
f_2	Grupo 4	10.592*	108.56*	2.031	Grupo 16	20.978*		
f_3		0.577	94.932*	1.563		17.551*		
f_1		2.557	4.596*	0.192		0.782	146.913*	1.234
f_2	Grupo 5	0.320	1.382	7.542*	Grupo 17	1.918	189.737*	0.2269
f_3		6.604*	16.921*	23.578*		2.438	146.074*	0.047
f_1		2.018	3.854*	0.579		2.041	126.32*	1.021
f_2	Grupo 6	6.651*	64.896*	1.992	Grupo 18	3.500	145.611*	1.185
f_3		6.841*	116.79*	2.099		3.244	102.327*	0.851
f_1		5.275*	178.65*	2.65		17.511*	149.915*	2.669
f_2	Grupo 8	3.439	295.72*	2.454	Grupo 19	16.929*	243.585*	2.898
f_3		0.995	299.78*	1.664		10.082*	244.302*	1.888
f_1		0.626				0.106	19.967*	6.832*
f_2	Grupo 10	0.088			Grupo 20	0.101	8.329*	10.392*
f_3		0.076				0.223	0.885	8.684*
f_1		0.291	357.14*	1.771		1.466	110.05*	1.881
f_2	Grupo 11	1.432	651.77*	4.006*	Grupo 21	1.872	215.285*	1.317
f_3		2.453	703.26*	4.927*		1.509	228.367*	0.567
f_1		0.231	137.841*	0.900		2.549		
f_2	Grupo 13	0.033	165.42*	0.105	Grupo 22	2.614		
f_3		0.004	299.348*	0.0421		2.151		

Vamos proceder à análise dos resultados apresentados na Tabela 3.20, considerando sempre que o nível de significância das nossas conclusões é de 5%. Relativamente à hipótese de ausência de efeitos do factor cultivar verificamos que:

1. para qualquer das três funções analisadas, essa hipótese não é rejeitada para os grupos 10, 11, 13, 17, 18, 20, 21 e 22. Estes grupos são compostos por cultivares com um comportamento bastante homogéneo;
2. para qualquer das três funções analisadas, essa hipótese é rejeitada para os grupos 3, 14, 16 e 19. Estes grupos são compostos por cultivares com um comportamento não homogéneo;

3. Para duas das três funções consideradas a mesma hipótese é rejeitada para os grupos 4 e 6;
4. Se considerarmos que a função que nos dá informação mais relevante é a f_3 , pois é a que corresponde ao ponto médio do intervalo de variação para os índices ambientais, a mesma hipótese é rejeitada para os grupos 3, 5, 6, 14, 16 e 19.

Relativamente à hipótese de ausência de efeitos do factor ano, verificamos que a mesma é sempre rejeitada, excepto para o grupo 5 com a função f_2 e para o grupo 20 com a função f_3 .

Relativamente à hipótese de ausência de interacções entre as combinações de níveis dos factores cultivar e ano verificamos que:

1. essa hipótese não é rejeitada para qualquer das funções consideradas para os grupos 6, 8, 13, 17, 18, 19 e 21;
2. é rejeitada para qualquer das funções utilizadas para o grupo 20;
3. é rejeitada para duas das três funções consideradas para os grupos 5 e 11.

O ideal seria a existência de homogeneidade no comportamento dos cultivares afectos a cada um dos grupos, pelo que esta análise nos permite concluir que existiram algumas falhas na condução deste melhoramento. A hipótese de ausência de efeitos para o factor cultivar é rejeitada para cerca de 50% dos grupos. No que respeita à grande heterogeneidade que se verifica relativamente aos anos, deverá em grande parte, dever-se a questões de ordem climática.

Podemos completar a nossa análise, através da aplicação de métodos de comparação múltipla, utilizando os métodos de Scheffé e de Bonferroni que nos permitirão identificar, dentro dos grupos, para os quais se rejeitaram as hipóteses, quais foram os cultivares e/ou os anos responsáveis pela rejeição das hipóteses de nulidade. Apresentaremos primeiramente as tabelas com os resultados dos métodos, e faremos posteriormente alguns comentários. Assinalamos com o símbolo (*), os casos em que existem diferenças significativas ao nível de 5% pelo método de Scheffé e com o símbolo (') os casos em que as mesmas se verificam pelo método de Bonferroni. As tabelas que apresentam os resultados dos métodos de comparação múltipla para os cultivares, contêm nas duas primeiras colunas as diferenças mínimas significativas para os métodos de Bonferroni e Scheffé respectivamente e nas colunas seguintes as diferenças em valor absoluto para os pares de observações. As tabelas que apresentam os resultados dos métodos para os anos, têm para cada par de anos uma coluna em que

Tabela 3.21: Método de Scheffé e Bonferroni para os cultivares do Grupo 3

função	Bonferroni	Scheffé	(1,2)	(1,3)	(2,3)
f_1	0.440	0.566	1.377*'	0.764*'	0.612*'
f_2	0.286	0.367	0.983*'	0.470*'	0.513*'
f_3	0.265	0.340	0.787*'	0.323'	0.463*'

se indica a diferença observada em valor absoluto entre os pares de observações/diferença mínima significativa pelo método de Scheffé.

Tabela 3.22: Método de Scheffé e Bonferroni para os cultivares do Grupo 4

função	Bonferroni	Scheffé	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
f_1	1.056	1.543	0.762	2.798*'	1.367'	2.036*'	2.13*'	4.165*'
f_2	0.556	0.816	0.362	1.144*'	0.441	0.783'	0.803'	1.586*'
f_3	0.520	0.765	0.161	0.317	0.022	0.156	0.139	0.296

Tabela 3.23: Método de Scheffé para os anos do Grupo 4

função	(1,2)	(1,3)	(2,3)
f_1	5.124/1.687*	7.754/1.203*	2.629/1.735*
f_2	0.181/0.866	4.065/0.733*	3.883/0.869*
f_3	2.291/0.812*	2.219/0.685*	4.510/0.814*

Tabela 3.24: Método de Scheffé para os anos do Grupo 8

função	(1,2)	(1,3)	(2,3)
f_1	3.513/0.936*	3.417/0.862*	6.929/0.902*
f_2	2.623/0.542*	2.906/0.537*	5.530/0.558*
f_3	2.178/0.484*	2.651/0.487*	4.829/0.484*

Tabela 3.25: Método de Scheffé e Bonferroni para os cultivares do Grupo 14

função	Bonferroni	Scheffé	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
f_1	0.660	0.969	0.585	0.579	0.685'	1.164*'	0.100	1.264*'
f_2	0.396	0.581	0.419'	0.347	0.645*'	0.767*'	0.225	0.992*'
f_3	0.312	0.458	0.337'	0.230	0.626*'	0.568*'	0.288	0.856*'

Tabela 3.26: Método de Scheffé e Bonferroni para os cultivares do Grupo 16

função	Bonferroni	Scheffé	(1,2)	(1,3)	(2,3)
f_1	0.567	0.728	1.682*'	0.347	1.336*'
f_2	0.350	0.450	1.046*'	0.031	1.016*'
f_3	0.298	0.383	0.753*'	0.093	0.846*'

Tabela 3.27: Método de Scheffé para os anos do Grupo 18

função	(1,2)	(1,3)	(2,3)
f_1	6.676/1.069*	4.309/0.953*	2.458/0.996*
f_2	4.406/0.655*	3.035/0.628*	1.371/0.652*
f_3	3.225/0.583*	2.398/0.597*	0.827/0.636*

Tabela 3.28: Método de Scheffé para os anos do Grupo 19

função	(1,2)	(1,3)	(2,3)
f_1	1.912/0.996*	7.425/1.105*	5.513/0.993*
f_2	0.664/0.652*	5.771/0.670*	5.107/0.674*
f_3	0.039/0.636	4.943/0.625*	4.904/0.638*

Após uma observação cuidada das tabelas apresentadas, podemos ao nível de significância de 5%, constatar que:

1. no Grupo 3 todos os pares de cultivares apresentam diferenças significativas;
2. no Grupo 4 com a função f_1 , todos os cultivares apresentam diferenças significativas excepto o par (flycatcher, te8501). Note-se que os cultivares te8504 e te8502 foram dominantes no contorno superior para a Opção *II*, justificando-se assim terem comportamentos distintos dos demais cultivares do grupo;
3. ao utilizar a função f_3 , estamos a analisar no ponto médio do intervalo de variação do índice ambiental, tendo-se obtido resultados bastante mais homogéneos no comportamento dos cultivares;
4. o Grupo 14 apresenta, para qualquer das funções usadas, diferenças significativas para todos os pares excepto para o par (mondego, te9102);
5. o Grupo 19 é constituído pelos cultivares te9203 e te9406. Ambos são cultivares dominantes no contorno superior obtido pela Opção *I*, mas o cultivar te9406 é dominante à direita pelo que se justifica perfeitamente a hipótese de homogeneidade ter sido rejeitada.

Capítulo 4

Índices Ambientais

4.1 Considerações Prévias

Neste capítulo construiremos um modelo para os índices ambientais. As variáveis controladas serão variáveis dummy correspondentes a:

- Anos;
- Locais;
- Grupos de cultivares.

Os coeficientes ajustados medirão os efeitos dessas variáveis. Em particular utilizaremos os coeficientes correspondentes a anos para tentar explicar as variações observadas nos máximos das funções objectivo apresentadas na tabela 3.17 do terceiro capítulo.

Na secção 4.2 deste capítulo apresentaremos os resultados teóricos que nos serão úteis para discutirmos, na secção 4.3, o ajustamento do modelo e finalmente, na última secção, consideraremos as funções objectivo.

4.2 Modelo

Observe-se que, dado haver grupos de cultivares que aparecem nos mesmos pares (local, ano), não se poderá analisar separadamente a acção dos cultivares que os compõem, o que para o nosso objectivo não causa qualquer problema. Assim, teremos de considerar variáveis dummy para esses grupos e não para os cultivares que os constituem. Este facto resulta possivelmente de se ter trabalhado, nos vários anos, com blocos completos. Se se tivesse optado por blocos incompletos teria havido mais possibilidades de desencontrar os pares (local, ano) em que os vários cultivares foram utilizados. Esta possibilidade surge, por exemplo, quando se utiliza delineamentos α , ver John and Williams (1995).

Sendo \mathbf{V}_b o vector dos índices ambientais, consideramos que o mesmo pode ser estimado com base no:

- ajustamento global utilizando os dados de todos os anos, que temos designado por Opção *I*;
- ajustamento fraccionado realizado ano a ano, que temos designado por Opção *II*.

Como os pares (local, ano) são os mesmos para as duas Opções, podemos emparelhar os resultados obtidos, o que permitirá pôr, uma vez mais em evidência, a estabilidade dos resultados obtidos pela aplicação da ACR neste contexto. Com efeito, seguiremos para as duas Opções tratamentos paralelos, tomando como variáveis explicativas, além duma variável constantemente igual a 1, variáveis "dummy" associadas a anos, locais e cultivares.

Sendo u_1, \dots, u_q , as variáveis dummy associadas a anos[locais], tem-se evidentemente:

$$\sum_{j=1}^q u_j = 1 \quad (4.2.1)$$

visto que cada par (local,ano) corresponderá a uma e apenas uma das componentes do vector de índices ambientais. Como temos uma variável constantemente igual a 1, para diminuir as possibilidades de existência de multicolinearidade, eliminamos u_q do modelo. Nas linhas da matriz do modelo em que u_q tomaria valor 1, as u_1, \dots, u_{q-1} tomam o valor -1 , uma vez que sendo $\delta_1, \dots, \delta_q$ os coeficientes para os anos[locais], se admite

$$\sum_{j=1}^q \delta_j = 0, \quad (4.2.2)$$

ou seja,

$$\delta_q \cdot 1 = \sum_{j=1}^{q-1} \delta_j \cdot (-1). \quad (4.2.3)$$

O cultivar Anza aparece em todos os pares (local, ano). A este cultivar teria de se associar uma variável dummy constantemente igual a 1. Daqui resultaria que o efeito dessa variável se confundia com o termo independente, por esta razão não a consideramos no modelo. Se tivermos p_1 , p_2 e p_3 variáveis dummy para os anos, locais e grupos de cultivares respectivamente, a matriz do modelo terá

$$h = 1 + p_1 + p_2 + p_3 \quad (4.2.4)$$

colunas. A primeira destas será constituída por elementos iguais a 1. Passando ao segundo grupo de colunas, nas linhas correspondentes ao i -ésimo ano, $i \leq p_1$, os elementos serão nulos salvo os da i -ésima coluna dentro do grupo que serão iguais a 1, essa coluna terá índice $i + 1$.

Se a linha corresponder ao último ano, os elementos situados nestas colunas serão como referimos iguais a -1 . Da mesma maneira se constroem as colunas associadas aos locais. Nas colunas correspondentes aos grupos de cultivares, os elementos de \mathbf{X} serão iguais a $1[0]$ nas linhas correspondentes aos pares (local, ano) em que esse grupo está presente[ausente]. Ter-se-á assim o modelo

$$\mathbf{V}_b = \mathbf{X}\boldsymbol{\delta}_h + \mathbf{e}_b \quad (4.2.5)$$

onde

1. \mathbf{V}_b é o vector das observações para os índices ambientais;
2. $\boldsymbol{\delta}_h$ é o vector dos coeficientes;
3. \mathbf{e}_b é o vector dos erros.

Admitamos ter $\mathbf{e}_b \sim N(\mathbf{0}_b, \sigma^2 \mathbf{I}_b)$, isto é, que o vector dos erros é normal com vector médio nulo e matriz de covariância $\sigma^2 \mathbf{I}_b$. Observe-se que os índices ambientais obtidos pelas Opções *I* e *II* tinham, como se viu na secção 3.2, Tabela 3.5, uma correlação quase unitária com os índices clássicos. Assim, é de admitir que os índices ambientais obtidos pela Opção *II* também serão homocedásticos. O coeficiente de correlação entre estes e os índices obtidos pela Opção *I*, leva-nos a estender esse pressuposto aos índices estimados pelo ajustamento global. Observe-se ainda que, anualmente, temos os cultivares considerados presentes em todos os locais, estando-se assim no caso equilibrado em que, ver Scheffé (1959) e Ito (1980), a ANOVA e técnicas relacionadas são robustas para a heterocedasticidade e ainda mais para a não normalidade. Assim o pressuposto da homocedasticidade para \mathbf{e}_b afigura-se como perfeitamente aceitável.

Um problema que se põe naturalmente é o da multicolinearidade. Com efeito, há por vezes associações entre locais e grupos de cultivares, sobretudo se os locais e os cultivares variarem com os anos. Com efeito, ao longo do programa de melhoramento os cultivares vão entrando e saindo, pelo que se os locais forem variando, surgem as associações atrás referidas. Para minimizar este problema, seria conveniente ter uma rede fixa de locais onde os ensaios seriam realizados, o que não se verifica no plano de melhoramento que temos estado a analisar.

Na aplicação que temos vindo a tratar, teremos de realizar o ajustamento na presença de multicolinearidade. Para o efeito, podemos recorrer a uma técnica baseada na análise espectral da matriz $\mathbf{X}'\mathbf{X}$, ver Judge, G.G et al (1988, pág 865). Assim, sendo $\lambda_1, \dots, \lambda_g$, com $g < h$, os valores próprios não nulos da matriz $\mathbf{X}'\mathbf{X}$ e $\boldsymbol{\alpha}_{1,h}, \dots, \boldsymbol{\alpha}_{g,h}$ os correspondentes

vectores próprios, substitui-se como matriz do modelo, \mathbf{X} por

$$\dot{\mathbf{X}} = \mathbf{X}\mathbf{K} = \begin{bmatrix} Z_{1,1} & \cdots & Z_{1,j} & \cdots & Z_{1,g} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{b,1} & \cdots & Z_{b,j} & \cdots & Z_{b,g} \end{bmatrix} = \left[\mathbf{Z}_{1,b}, \dots, \mathbf{Z}_{g,b} \right] \quad (4.2.6)$$

onde $\mathbf{K} = [\boldsymbol{\alpha}_{1,h}, \dots, \boldsymbol{\alpha}_{g,h}]$.

Convém talvez referir que as colunas de $\dot{\mathbf{X}}$ contêm os valores das chamadas componentes principais. Estas serão as combinações lineares das variáveis explicativas iniciais, obtidas utilizando como vectores de coeficientes, os vectores próprios de $\mathbf{X}'\mathbf{X}$ associados a valores próprios não nulos.

As matrizes \mathbf{X} e $\dot{\mathbf{X}}$ terão o mesmo espaço imagem, o que permite rescrever o modelo como,

$$\mathbf{V}_b = \dot{\mathbf{X}}\boldsymbol{\gamma}_g + \mathbf{e}_b \quad (4.2.7)$$

e como $Car(\dot{\mathbf{X}}) = g$, eliminou-se a multicolinearidade já que os vectores coluna de $\dot{\mathbf{X}}$ são linearmente independentes. Obtém-se agora o estimador centrado de mínimos quadrados

$$\tilde{\boldsymbol{\gamma}}_g = (\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1}\dot{\mathbf{X}}'\mathbf{V}_b, \quad (4.2.8)$$

normalmente distribuído e com matriz de covariância

$$Cov(\tilde{\boldsymbol{\gamma}}_g) = \sigma^2(\dot{\mathbf{X}}'\dot{\mathbf{X}})^{-1} = \sigma^2(\mathbf{K}'\mathbf{X}'\mathbf{X}\mathbf{K})^{-1} = \sigma^2\mathbf{diag}(\lambda_1, \dots, \lambda_g)^{-1}, \quad (4.2.9)$$

independente da soma dos quadrados dos resíduos

$$S_{cp} = \mathbf{V}_b'\mathbf{V}_b - \mathbf{V}_b'\dot{\mathbf{X}}\tilde{\boldsymbol{\gamma}}_g, \quad (4.2.10)$$

que se distribui como o produto de σ^2 por um qui-quadrado central com $b - g$ graus de liberdade, pondo-se $S_{cp} \sim \sigma^2\chi_{b-g}^2$.

O vector ajustado será,

$$\tilde{\mathbf{V}}_b = \sum_{j=1}^g (\tilde{\gamma}_j \cdot \mathbf{Z}_{j,b}). \quad (4.2.11)$$

Sendo $\mathbf{x}_{1,b}, \dots, \mathbf{x}_{h,b}$ os vectores coluna da matriz \mathbf{X} e $\alpha_{1,j}, \dots, \alpha_{h,j}$ as componentes de $\boldsymbol{\alpha}_{j,h}$, $j = 1, \dots, g$, vem, atendendo a (4.2.6)

$$\mathbf{Z}_{j,b} = \sum_{l=1}^h \alpha_{l,j}\mathbf{x}_{l,b}, j = 1, \dots, g \quad (4.2.12)$$

logo, substituindo em (4.2.11), obtém-se:

$$\tilde{\mathbf{V}}_b = \sum_{j=1}^g (\tilde{\gamma}_j \cdot \mathbf{Z}_{j,b}) = \sum_{j=1}^g (\tilde{\gamma}_j \cdot \sum_{l=1}^h \alpha_{l,j}\mathbf{x}_{l,b}) = \sum_{l=1}^h (\sum_{j=1}^g \alpha_{l,j}\tilde{\gamma}_j)\mathbf{x}_{l,b} = \sum_{l=1}^h (\tilde{\delta}_l\mathbf{x}_{l,b}) \quad (4.2.13)$$

sendo $\tilde{\delta}_l, l = 1, \dots, h$, as componentes do vector que inicialmente se pretendia estimar. Tem-se agora o vector de estimadores

$$\tilde{\delta}_h = \mathbf{K} \cdot \tilde{\gamma}_g \quad (4.2.14)$$

o qual, ver Seber (1980, pág. 5), será normal, com vector médio $\delta_h = \mathbf{K} \cdot \gamma_g$ e matriz de covariância

$$Cov(\tilde{\delta}_h) = \sigma^2 \mathbf{W} \quad (4.2.15)$$

onde $\mathbf{W} = \mathbf{K} \mathbf{D} \mathbf{K}' = [w_{l,j}]$, com $\mathbf{D} = \mathbf{diag}(\lambda_1^{-1}, \dots, \lambda_g^{-1})$, sendo independente de $S_{cp} \sim \sigma^2 \chi_{b-g}^2$. Assim, para testar

$$H_{l,o} : \delta_l = \delta_{l,o}, l = 1, \dots, h \quad (4.2.16)$$

ter-se-á a estatística de teste

$$t(\delta_{l,o}) = \frac{\tilde{\delta}_l - \delta_{l,o}}{\sqrt{w_{l,l} \frac{S_{cp}}{b-g}}}, l = 1, \dots, h \quad (4.2.17)$$

que, quando a hipótese testada se verifica, tem distribuição t central com $b - g$ graus de liberdade.

Para decidir o número de componentes principais a utilizar, podemos utilizar um critério usual e muito simples. Estando os valores próprios da matriz $\mathbf{X}' \mathbf{X}$ ordenados por ordem decrescente, o número de componentes principais a utilizar será igual ao número de valores próprios superiores a

$$\frac{tr(\mathbf{D}^{-1})}{g} = \frac{\sum_{j=1}^g \lambda_j}{g}. \quad (4.2.18)$$

Na aplicação que temos vindo a utilizar, temos para cada par (local, ano) quatro repetições do índice ambiental estimado. Podemos trabalhar com as médias dessas quatro observações, sem que isso obrigue a obter um modelo diferente. Vejamos que, quando há grupos de r linhas idênticas na matriz \mathbf{X} do modelo, temos

$$\mathbf{X} = \mathbf{X} \otimes \mathbf{1}_r. \quad (4.2.19)$$

Podemos aligeirar o modelo, substituindo \mathbf{X} por \mathbf{X} . e o vector \mathbf{Y} das observações pelo vector \mathbf{Y} . das m médias dos grupos de r observações da variável dependente consecutivas. Com efeito, sendo $\tilde{\gamma}$ um vector de coeficientes estimado tem-se,

$$\mathbf{X} \tilde{\gamma} = (\mathbf{X} \cdot \tilde{\gamma}) \otimes \mathbf{1}_r, \quad (4.2.20)$$

vindo

$$\mathbf{Y} - \mathbf{X} \tilde{\gamma} = \mathbf{Y} - \mathbf{Y} \otimes \mathbf{1}_r + \mathbf{Y} \otimes \mathbf{1}_r - (\mathbf{X} \cdot \tilde{\gamma}) \otimes \mathbf{1}_r = \mathbf{Y} - \mathbf{Y} \otimes \mathbf{1}_r + (\mathbf{Y} - \mathbf{X} \cdot \tilde{\gamma}) \otimes \mathbf{1}_r. \quad (4.2.21)$$

Ora $\mathbf{Y} - \mathbf{Y} \cdot \mathbf{1}_r$ pertence ao espaço de nulidade de $I_m \otimes \mathbf{1}'_r$, que representamos por $N(I_m \otimes \mathbf{1}'_r)$ que é, ver Seber (1980, pág. 2) o complemento ortogonal do espaço imagem $R(I_m \otimes \mathbf{1}_r)$ de $I_m \otimes \mathbf{1}_r$. Assim $\mathbf{Y} - \mathbf{Y} \cdot \mathbf{1}_r$ será ortogonal a $(\mathbf{Y} - \mathbf{X} \cdot \tilde{\boldsymbol{\gamma}}) \otimes \mathbf{1}_r$ que pertence a $R(I_m \otimes \mathbf{1}_r)$, tendo-se pois

$$\|\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\gamma}}\|^2 = \|\mathbf{Y} - \mathbf{Y} \cdot \mathbf{1}_r\|^2 + r \|\mathbf{Y} \cdot - \mathbf{X} \cdot \tilde{\boldsymbol{\gamma}}\|^2 \quad (4.2.22)$$

o que mostra que, efectivamente, se pode trabalhar com o modelo compactado.

4.3 Ajustamento

Neste parágrafo vamos uma vez mais utilizar os dados de Melhoramento do Trigo Mole para proceder à modelação dos índices ambientais. Teremos pois dados de onze anos referentes a quinze locais e aos grupos de cultivares indicados na Tabela 3.18 da secção 3.10. Vamos proceder à estimação do modelo tal como se indicou na secção 4.1. Nas Tabelas 4.1 e 4.2 apresentam-se respectivamente os coeficientes ajustados bem como os testes t para as hipóteses de nulidade dadas por (4.2.15) para as Opções I e II .

Tabela 4.1: Coeficientes ajustados e testes t na Opção I

	Variáveis dummy	Coeficientes ajustados	Teste t
	ordenada na origem	1.275	134.519 ***
Anos	1986	0.282	14.276***
	1987	0.144	45.983***
	1988	0.142	7.989***
	1989	0.628	15.356***
	1990	-0.827	-15.356***
	1991	0.232	9.195***
	1992	0.225	10.274***
	1995	-0.234	-15.744***
	1997	-0.0058	-0.644(ns)
	1999	0.479	11.243***
	2000	-1.064	-1.399(ns)
Locais	Almeirim	0.082	30.527***
	Évora	0.223	55.430***
	Coruche	0.082	30.526***
	Mirandela	0.262	39.989***
	Comenda	0.171	23.992***
	Fundão	0.231	43.191***
	E.N.M.P.	0.134	26.838***
	Lamaçais	0.008	2.798*
	Beja	-0.108	-5.126***
	Benavila	-0.148	-20.521***
	Elvas	-0.185	-39.981***
	Revilheira	-0.127	-11.557***
	Santarém	0.049	12.890***
	Abrantes	0.049	12.890***
	V.F. Xira	0.239	19.377***
M.Alhos	-0.968	-0.589(ns)	
grupos de cultivares	2	0.716	65.615***
	3	0.301	15.975***
	4	0.626	55.865***
	5	0.465	26.808***
	6	0.324	13.803***
	7	0.163	11.823***
	8	-0.002	-0.015(ns)
	9	0.650	64.049***
	10	0.645	16.88***
	11	-0.161	-9.02***
	12	0.09	7.595***
	13	-0.557	-15.748***
	14	0.251	10.871***
	15	0.496	19.813***
	16	0.245	11.921***
	17	0.029	1.497(ns)
	18	0.043	2.032(ns)
	19	0.315	20.194***
	20	-0.202	-5.866***
21	0.530	22.291***	
22	0.517	17.825***	

Refira-se ainda que utilizamos neste ajustamento a decomposição da matriz $\mathbf{X}'\mathbf{X}$ em componentes principais e a regra prática que referimos atrás levou-nos a usar nove componentes principais. Verifica-se que, em quaisquer das Opções, a grande maioria dos coeficientes estimados são significativos, aos níveis de significância de * 5%, ** 1% e *** 0.1%, o que mostra o poder explicativo das variáveis consideradas.

Tabela 4.2: Coeficientes ajustados e testes t na Opção II

	Variáveis Dummy	Coeficientes ajustados	Teste t
Anos	ordenada na origem	1,284	134,733 ***
	1986	0,274	13,789***
	1987	0,137	43,557***
	1988	0,134	7,536***
	1989	0,580	14,117***
	1990	-0,808	-14,924***
	1991	0,202	7,992***
	1992	0,237	10,744***
	1995	-0,257	-17,126***
	1997	-0,027	-2,893**
	1999	0,485	11,324***
2000	-0,958	-1,255(ns)	
Locais	Almeirim	0,073	27,318***
	Évora	0,212	50,986***
	Coruche	0,073	27,318***
	Mirandela	0,255	37,916***
	Comenda	0,158	22,191***
	Fundão	0,215	39,891***
	E.N.M.P.	0,127	25,268***
	Lamaçais	-0,001	-0,217(ns)
	Beja	-0,143	-6,761**
	Benavila	-0,145	-20,009***
	Elvas	-0,189	-40,615***
	Revilheira	-0,129	-11,684***
	Santarém	0,034	8,932***
	Abrantes	0,034	8,932***
V.F. Xira	0,240	19,395***	
M.Alhos	-0,811	-0,491(ns)	
cultivares	2	0,698	63,631***
	3	0,304	16,006***
	4	0,635	56,346***
	5	0,471	26,992***
	6	0,331	14,008***
	7	0,162	11,667***
	8	-0,049	-0,400(ns)
	9	0,650	63,718***
	10	0,610	15,821***
	11	-0,169	-9,414***
	12	0,063	5,318***
	13	-0,546	-15,369***
	14	0,232	10,002***
	15	0,499	19,830***
16	0,266	12,921***	
17	0,039	2,017(ns)	
18	0,042	2,023(ns)	
19	0,320	20,397***	
20	-0,224	-6,467***	
21	0,547	22,861***	
22	0,544	18,649***	

4.4 Funções objectivo

Apresentámos na Tabela 3.17 da secção 3.9 os máximos das funções objectivo f_1, f_2, f_3 . Verificamos que os mesmos apresentam algumas oscilações ao longo dos anos. Surgiu assim a ideia de ver até que ponto, essas oscilações podem ser explicadas pelos coeficientes ajustados para os anos apresentados nas Tabelas 4.1 e 4.2. Para tal, procedemos ao ajustamento de três regressões lineares, uma por função, onde se consideraram os coeficientes dos anos como variável explicativa das oscilações dos valores máximos das funções f_1, f_2, f_3 . Apresentamos na Tabela 4.3 os valores obtidos para os coeficientes de regressão estimados, bem como o

valor da estatística de teste para a hipótese de nulidade dos coeficientes angulares das rectas de regressão ajustadas.

Tabela 4.3: Coeficientes ajustados para os máximos das funções objectivo

	f_1	f_2	f_3
$\tilde{\alpha}$	5.088	3.835	3.284
$\tilde{\beta}$	3.395	2.634	2.280
teste t para $\tilde{\beta}$	4.841	9.362	9.611

Observe-se que, os coeficientes angulares destas rectas de regressão ajustadas são significativamente diferentes de zero aos níveis de 1% para f_1 e de 0.1% para f_2 e f_3 . Portanto, podemos concluir a elevada significância do factor ano, nas oscilações dos máximos das funções objectivo.

Quanto aos coeficientes para os locais e grupos de cultivares, a variação dos locais de ano para ano e a não homogeneidade de grande parte dos grupos de cultivares, torna complicada a interpretação desses resultados.

Capítulo 5

Conclusões e Trabalhos Futuros

Julgamos ter mostrado que se pode:

1. validar a utilização da ACR para a interpretação conjunta dos resultados de planos de melhoramento;
2. utilizar a ACR como técnica de acompanhamento de planos de melhoramento;
3. modelar os índices ambientais de forma a quantificar-se o efeito nesses índices de cultivares ou grupos de cultivares, locais e anos.

Neste trabalho, realizou-se a modelação referida para estudar a influência da produtividade dos vários anos, nos máximos das funções objectivo. Consegue-se assim, quantificar a influência dos anos na evolução desses máximos.

No entanto, o aspecto mais relevante dos nossos resultados é o facto da ACR poder ser utilizada para acompanhar a realização de planos de melhoramento. Assim, esta técnica poderá passar a ser utilizada no fim de cada ano, para ajudar à tomada de decisão sobre os cultivares a manter no programa para o ano seguinte. Evitar-se-á desta forma, manter ou eliminar cultivares incorrectamente.

A ser assim, ter-se-á alargado a campo de aplicação da ACR, para além do seu domínio inicial. Este alargamento foi conseguido em duas vertentes:

1. aumento do prazo temporal, indo-se além dum número reduzido de anos (regra geral dois a três);
2. flexibilidade no conjunto de cultivares, na medida em que se passa apenas a "exigir" a existência de cultivares em comum em anos consecutivos.

Para além das duas vertentes de alargamento do âmbito de aplicação da técnica, pode salientar-se que se aprofundou a modelação dos índices ambientais, iniciada por Pereira (2004).

Apresentadas as principais conclusões do nosso trabalho, cabe-nos dizer algo sobre trabalhos futuros. Como vimos, o presente trabalho aponta para o alargamento do âmbito da aplicação da ACR. Esse alargamento revelou-se possível dada a robustez da técnica. Surge assim a ideia de tentar um alargamento posterior do campo de aplicação.

Num plano de melhoramento, há um fio conductor na escolha de cultivares, locais e tipos de delineamento. Assim, os ensaios que compõem um tal plano, foram delineados de acordo com esse fio conductor. Podemos por oposição a este caso, considerar conjuntos de ensaios sobre uma mesma cultura, delineados separadamente. Interessa ver em que medida é que a ACR é aplicável a tais conjuntos de ensaios. Uma primeira tentativa neste sentido foi feita com ensaios de videira, ver Pinto, I. e Mexia, J. T. (2004). Tencionamos agora desenvolver este campo de uma forma sistemática.

Apêndice A

Apresentamos seguidamente as tabelas de dados que serviram de base à realização dos testes para as hipóteses de ausência de efeitos e de interações realizados na secção 3.10. Note-se que os grupos 3, 10, 14, 16 e 22 apenas têm observações correspondentes aos anos de 1986, 1989, 1991, 1992 e 1999 respectivamente. Os grupos 1, 2, 7, 9, 12 e 15 são constituídos apenas por um cultivar não tendo sido considerados.

Grupo 3	f_1	f_2	f_3
te8401	6.567	4.594	3.608
sunbird	5.190	3.612	2.822
nelkant-s	5.802	4.124	3.285

Grupo 4	Anos	flycatcher	te8501	te8502	te8504
f_1	1986	6.537	7.139	6.719	6.187
	1987	4.927	5.222	7.036	4.303
	1988	4.727	4.593	5.204	4.334
f_2	1986	4.483	4.832	4.687	4.309
	1987	4.423	4.371	5.060	4.277
	1988	3.500	3.564	3.804	3.379
f_3	1986	3.456	3.679	3.656	3.370
	1987	4.170	3.946	4.072	4.264
	1988	2.887	3.050	3.103	2.900

Grupo 5	Anos	hanh-s	miwivet
f_1	1986	7.028	6.746
	1987	6.6130	6.117
f_2	1986	4.994	4.721
	1987	4.797	5.211
f_3	1986	3.976	3.708
	1987	3.889	4.758

Grupo 6	Anos	te8601	te8602
f_1	1986	5.124	5.670
	1987	4.823	4.988
f_2	1986	4.448	4.948
	1987	3.614	3.761
f_3	1986	4.109	4.587
	1987	3.009	3.147

Grupo 8	Anos	te8701	te8702
f_1	1988	3.874	4.391
	1989	5.570	6.208
	1990	2.485	2.363
f_2	1988	3.020	3.243
	1989	4.249	4.637
	1990	1.731	1.626
f_3	1988	2.593	2.669
	1989	3.589	3.852
	1990	1.354	1.257

Grupo 10	f_1	f_2	f_3
Liz1	6.300	4.905	4.207
Alva	6.522	4.953	4.169

Grupo 11	Anos	te8601	te8602
f_1	1989	6.318	5.974
	1990	2.596	2.742
f_2	1989	4.942	4.578
	1990	1.809	1.901
f_3	1989	4.253	3.880
	1990	1.416	1.480

Grupo 13	Anos	te8901	te8902
f_1	1990	3.054	3.153
	1991	5.720	5.420
f_2	1990	2.148	2.170
	1991	4.259	4.178
f_3	1990	1.694	1.679
	1991	3.528	3.557

Grupo 14	f_1	f_2	f_3
te9001	5.259	3.983	3.344
milan	4.674	3.563	3.007
te9003	5.839	4.329	3.575
te8906	4.574	3.337	2.719

Grupo 16	f_1	f_2	f_3
mondego	5.430	4.031	3.332
te9101	3.747	2.985	2.579
te9102	5.083	4.000	3.425

Grupo 17	Anos	te9111	te9112
f_1	1992	4.475	4.910
	1995	2.071	2.022
f_2	1990	3.496	3.748
	1991	1.719	1.844
f_3	1990	3.007	3.167
	1991	1.543	1.754

Grupo 18	Anos	te9113	te9114
f_1	1992	5.247	5.796
	1995	2.050	2.226
	1997	3.37	3.364
f_2	1992	3.939	4.355
	1995	1.864	2.023
	1997	2.616	2.643
f_3	1992	3.284	3.635
	1995	1.771	1.922
	1997	2.238	2.282

Grupo 19	Anos	te9203	te9406
f_1	1995	2.140	2.283
	1997	2.727	3.608
	1999	5.356	6.492
f_2	1995	2.023	2.106
	1997	2.104	2.689
	1999	4.588	5.312
f_3	1995	1.965	2.017
	1997	1.793	2.229
	1999	4.205	4.721

Grupo 20	Anos	te9203	te9406	te9303
f_1	1995	2.312	2.852	2.225
	1997	3.454	2.727	3.432
f_2	1995	2.032	2.479	1.864
	1997	2.579	2.104	2.633
f_3	1995	1.892	2.293	1.684
	1997	2.142	1.793	2.233

Grupo 21	Anos	te9503	te9504
f_1	1997	3.173	3.140
	1999	5.021	5.544
f_2	1997	2.323	2.353
	1999	4.186	4.533
f_3	1997	1.898	1.960
	1999	3.768	4.026

Grupo 22	f_1	f_2	f_3
te9712	5.395	4.624	4.238
te9713	6.047	5.077	4.591
te9714	5.338	4.708	4.393
te9715	5.160	4.457	4.106
te9716	5.199	4.541	4.213

Quando na secção 3.7 apresentámos os resultados dos contornos superiores na opção II, remetemos para o apêndice os gráficos dos contornos superiores para cada ano, os quais apresentamos seguidamente. Considerámos mais interessante apresentar os contornos superiores baseados nos ajustamentos sobre os índices ambientais reais. Assim em cada gráfico, no eixo das abcissas, observa-se a real variação dos índices ambientais ajustados para cada ano. Não consideramos o ano 2000 devido à falta de informação relativa ao mesmo. Devido ao número reduzido de observações para esse ano, o ajustamento realizado apresentou falta de qualidade tendo dado origem a um intervalo de variação para os índices ambientais estimado muito reduzido.

Figura A.1: Contorno Superior no ano de 1986

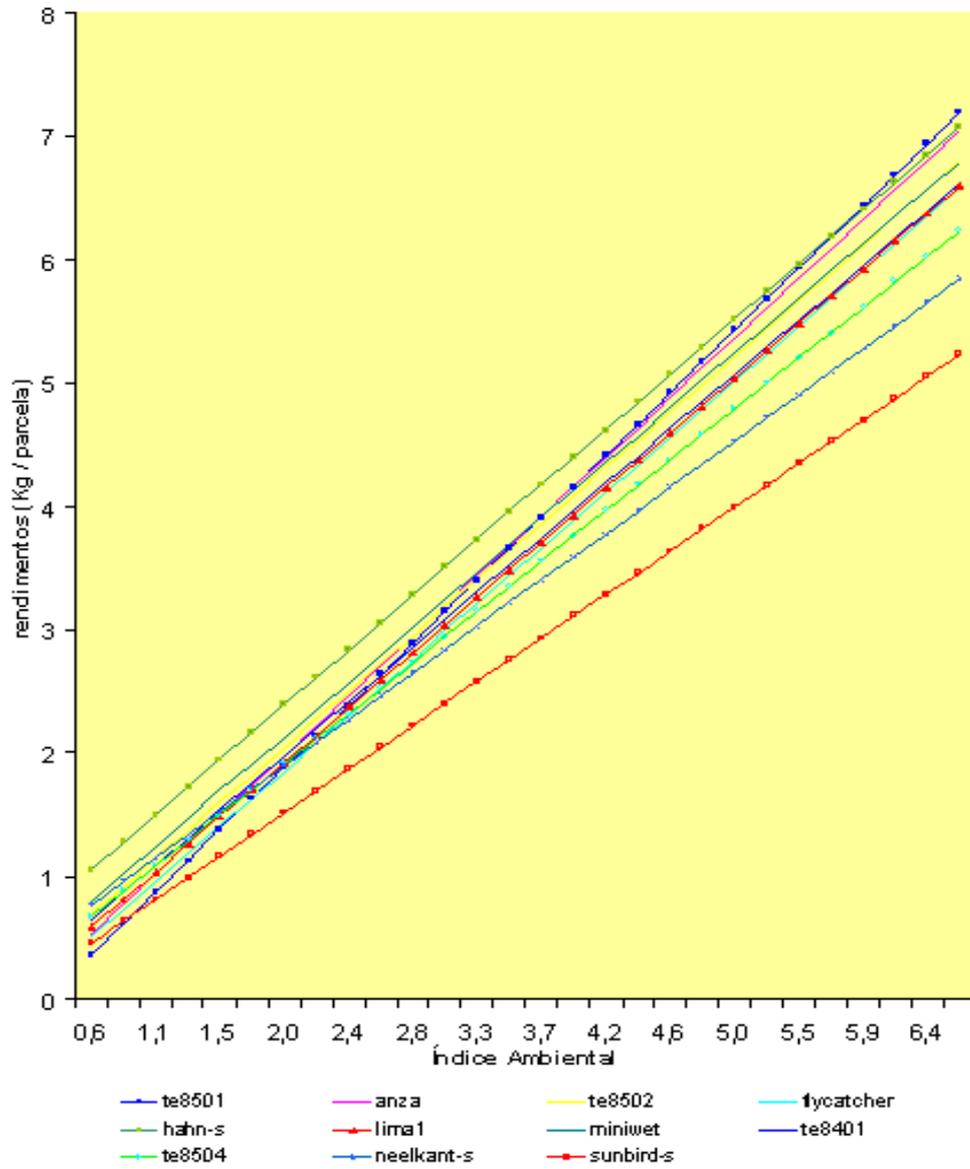


Figura A.3: Contorno Superior no ano de 1988

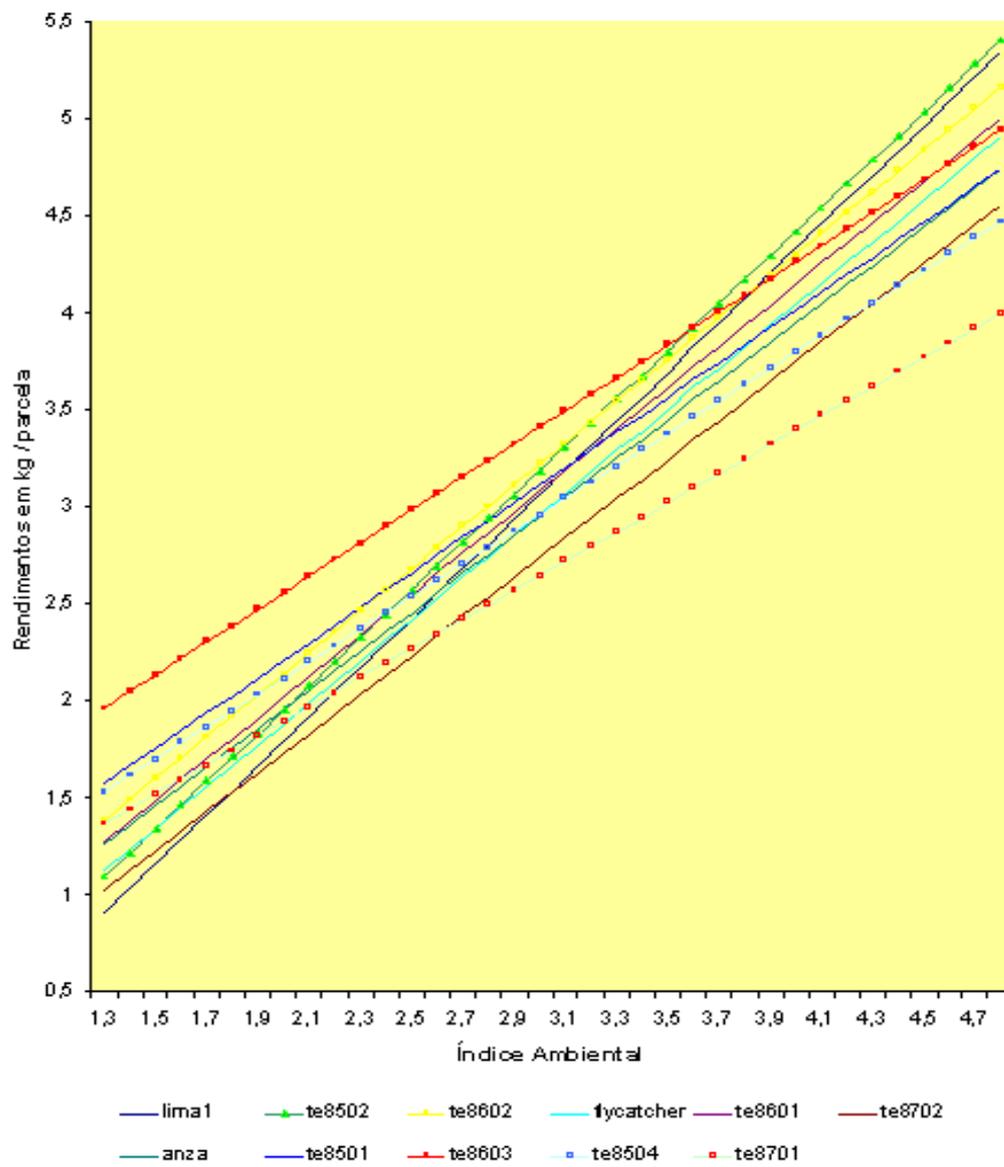


Figura A.4: Contorno Superior no ano de 1989

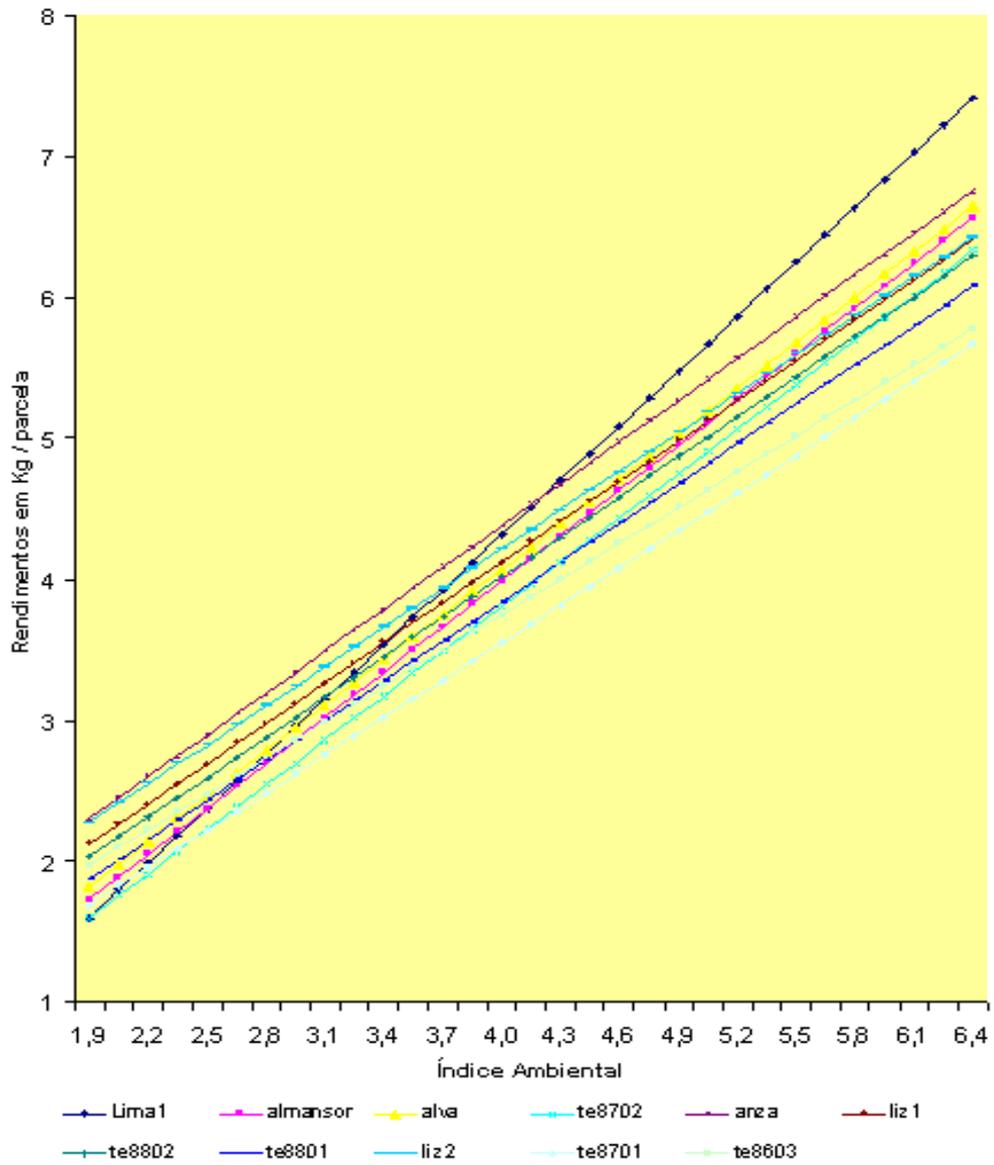


Figura A.6: Contorno Superior no ano de 1991

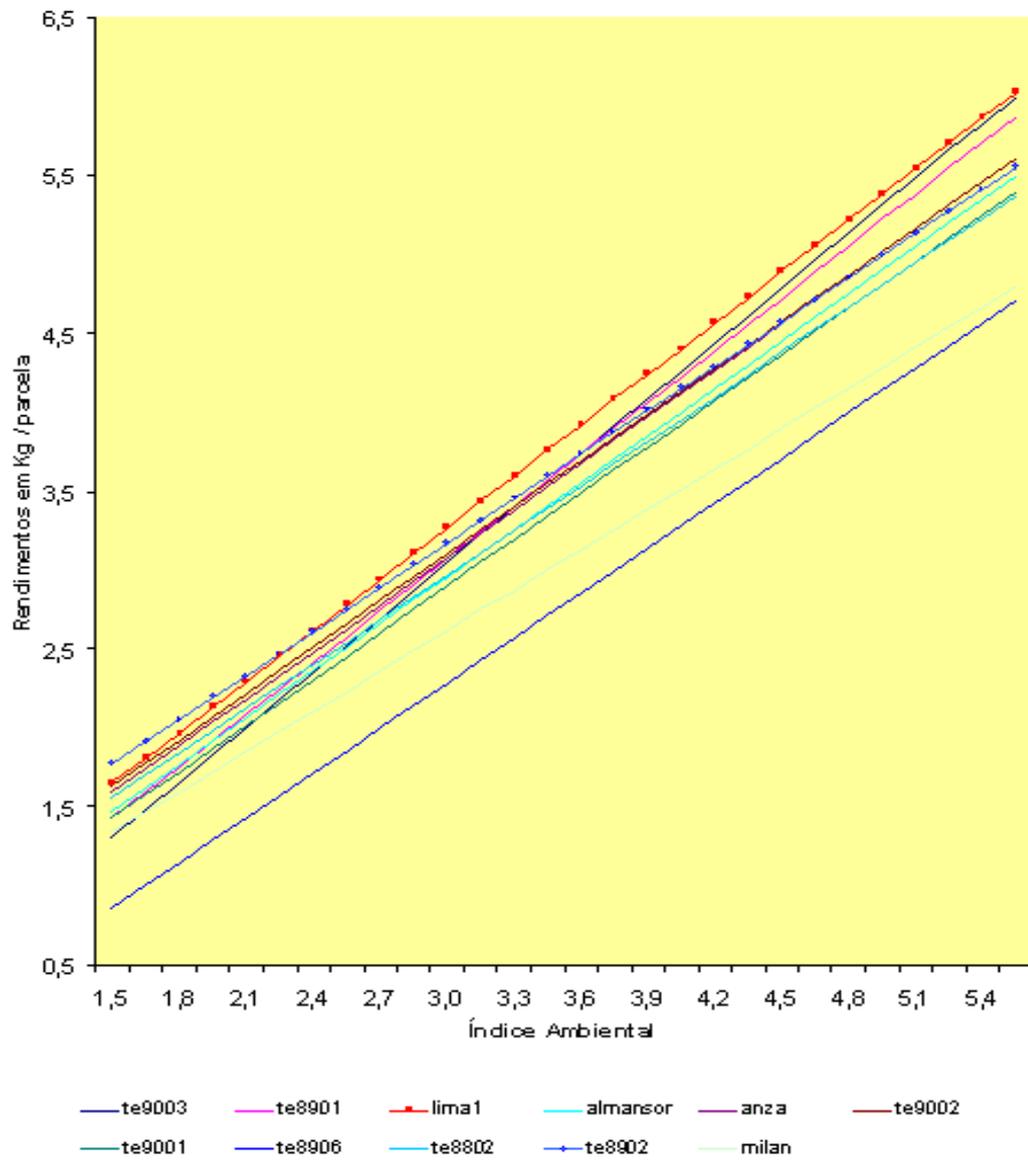


Figura A.7: Contorno Superior no ano de 1992

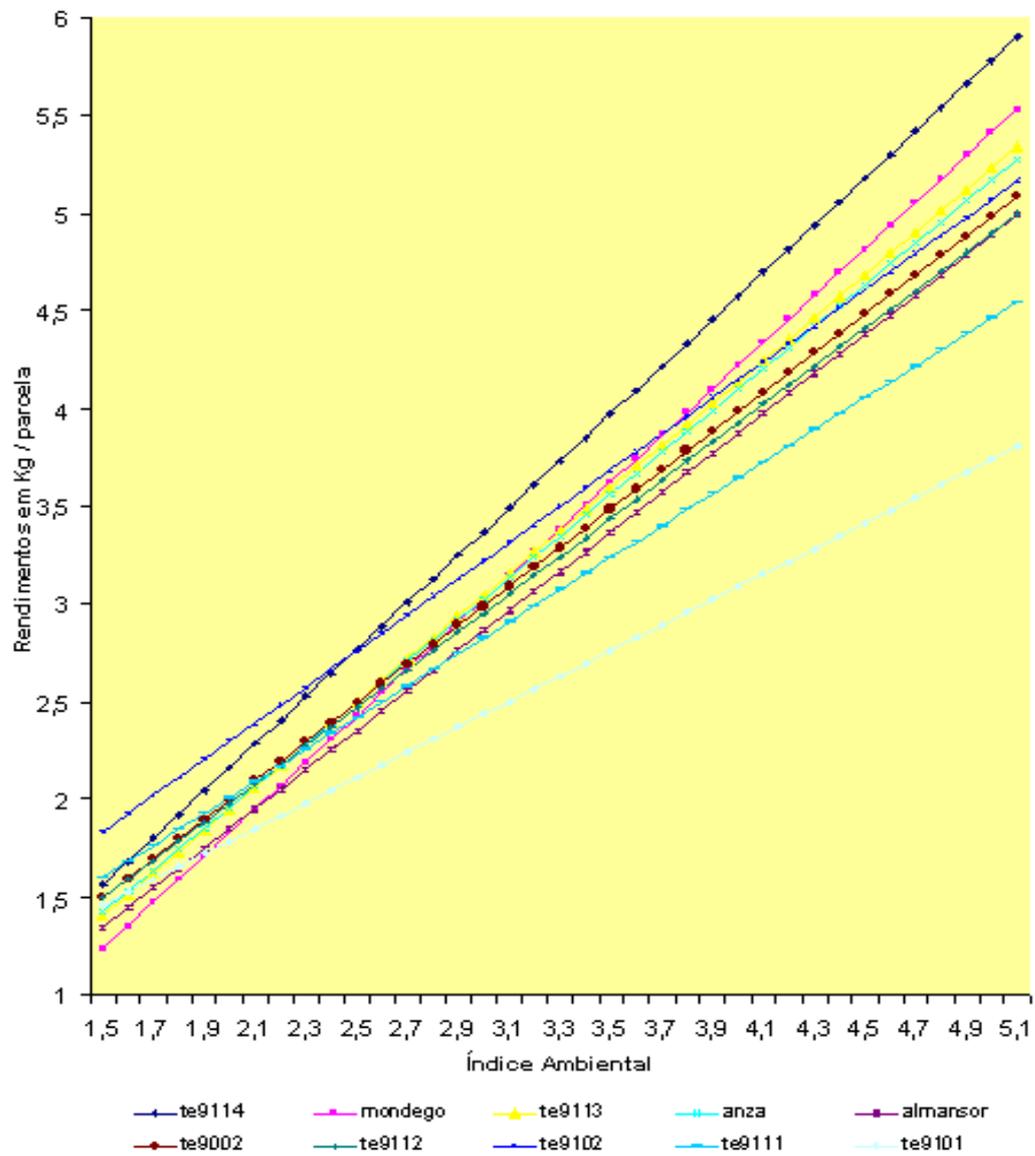


Figura A.8: Contorno Superior no ano de 1995

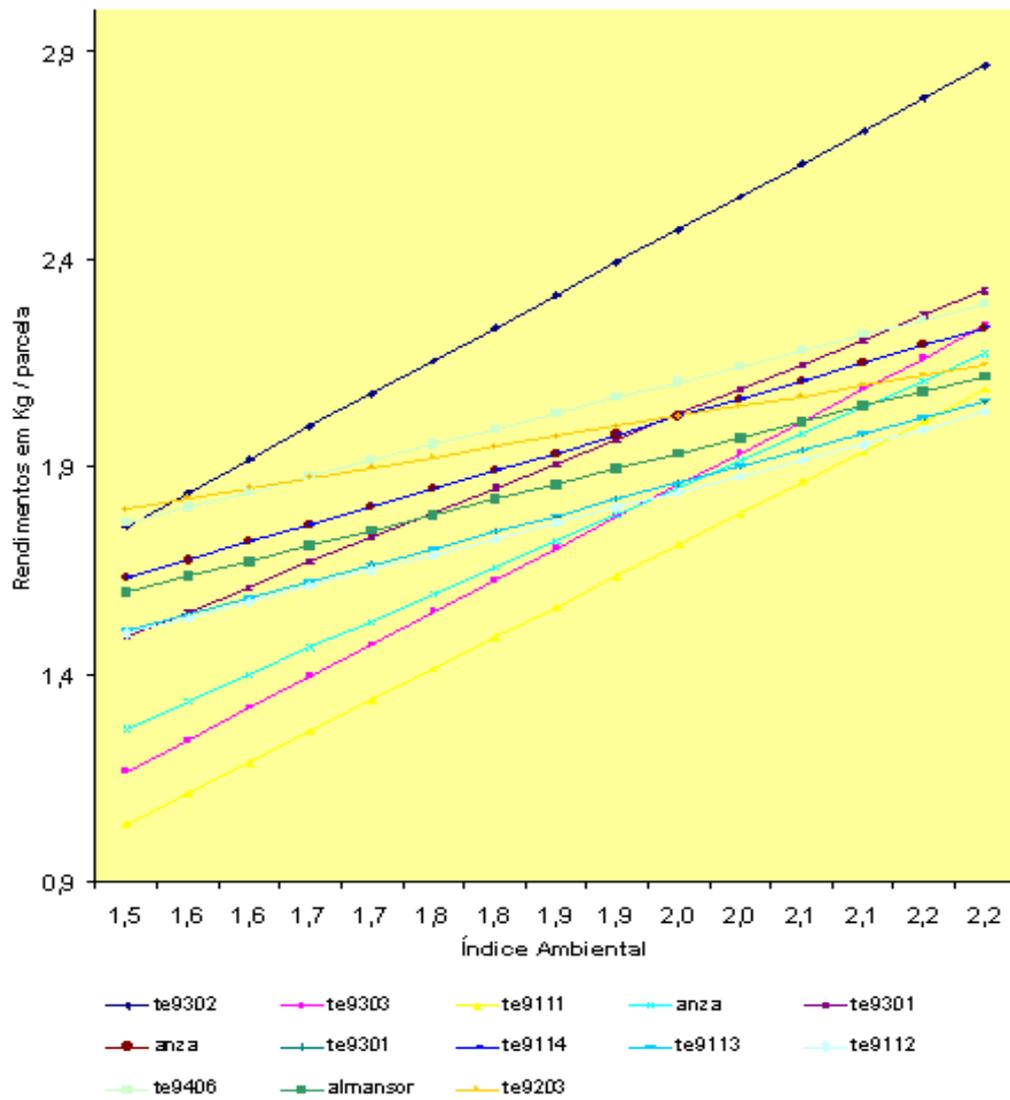


Figura A.9: Contorno Superior no ano de 1997

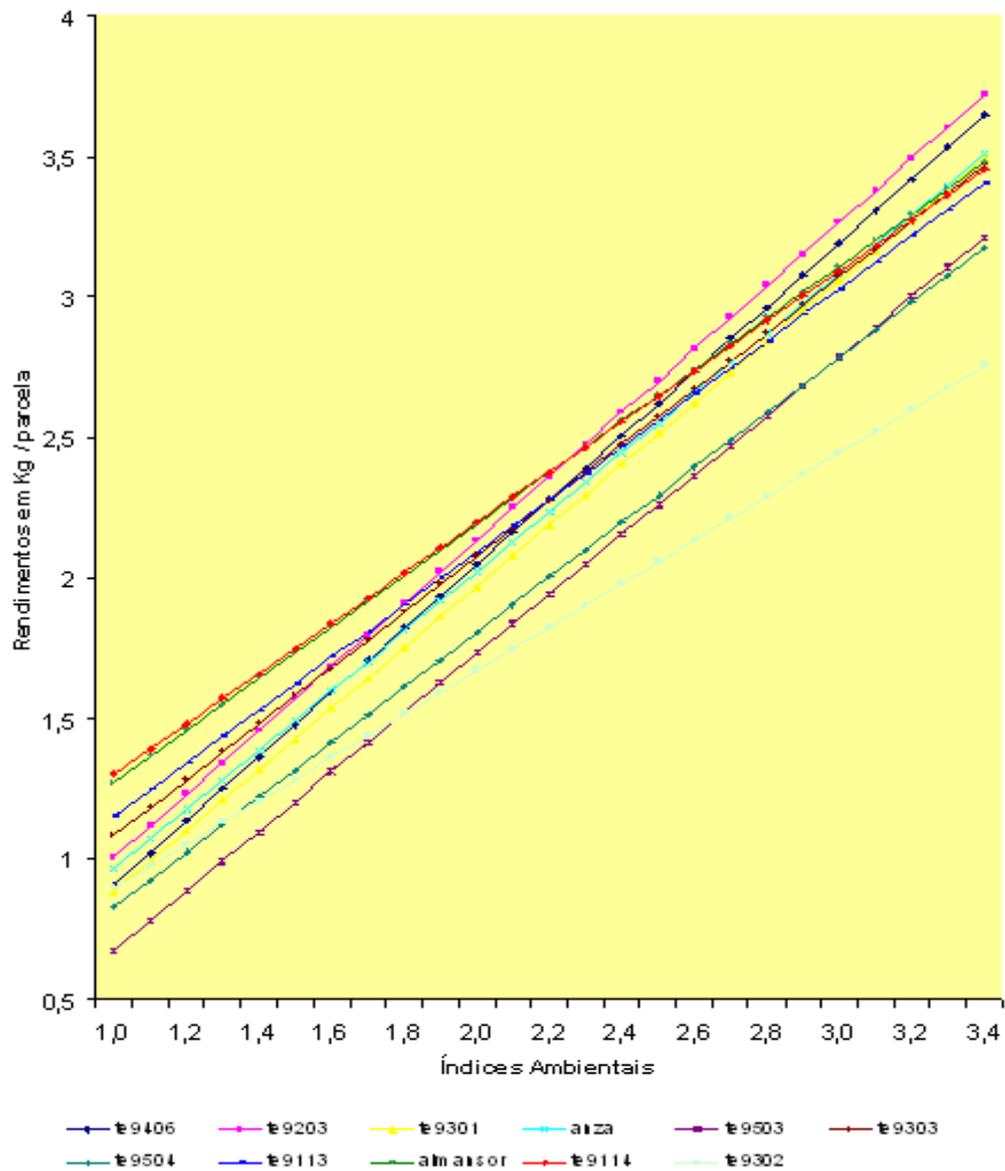
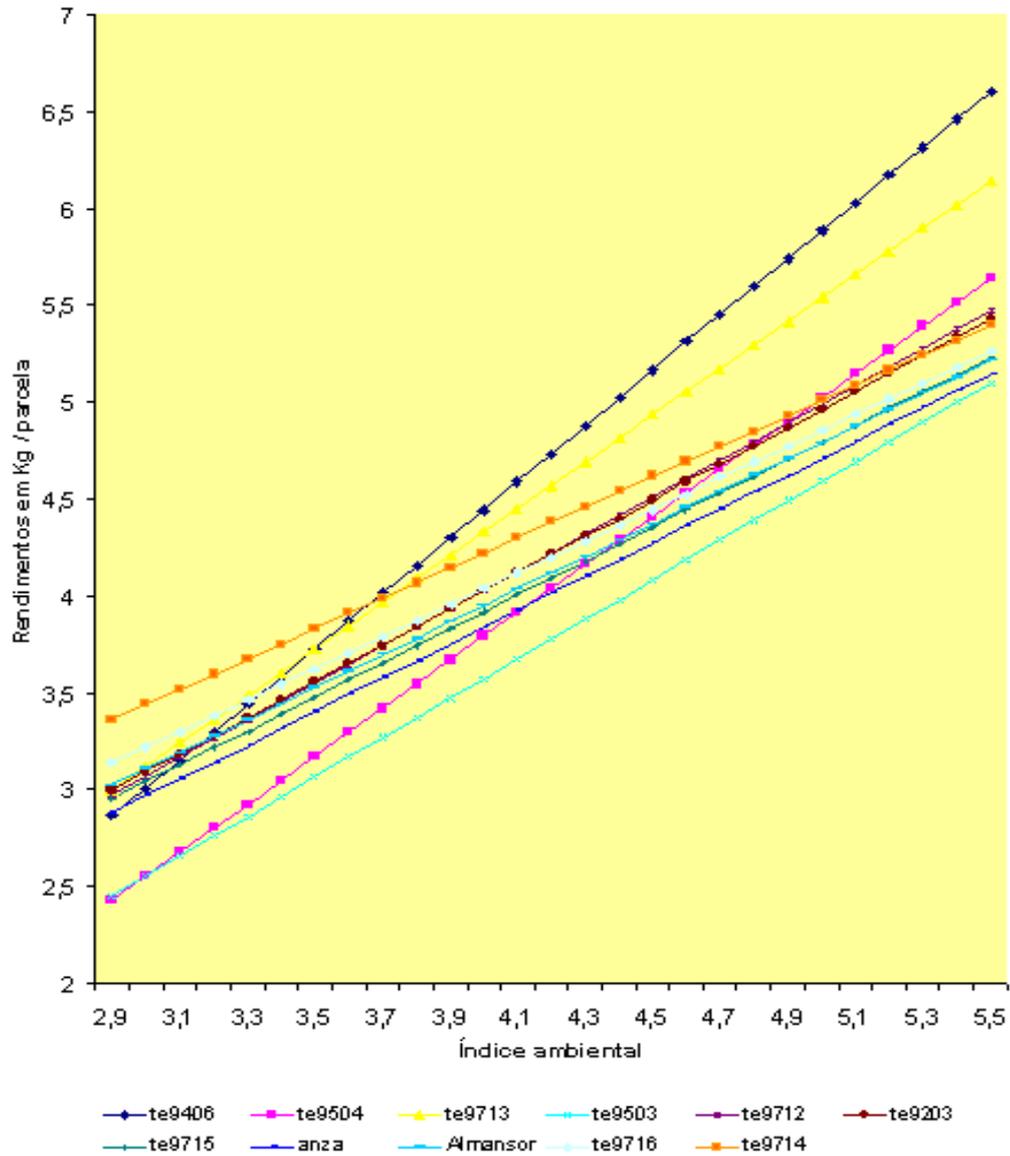


Figura A.10: Contorno Superior no ano de 1999



Bibliografia

- [1] Aastveit, A. H. and Mejza, S. (1992) - *A selected bibliography on statistical methods for the analysis of genotype \times environment interaction*. Biuletyn Oceny Odmian, 24 - 25 : 83 - 97.
- [2] Bazarra, M. S., Sherali, H. D. and Shetty, C. M. (1992) - *Nonlinear programming, Theory and Algorithms*. 2nd ed., Jonh Wiley & Sons - New York.
- [3] Becker, H. C. and J. Leon (1988) - *Stability analysis in plant breeding*. *Plant Breeding*, 101 : 1 - 23.
- [4] Bose, R. C. (1950) - *Least Squares aspects of analysis of variance*. Mimeo Séries 9, Institute of Statistics, University of North Carolina, Chapel Hill.
- [5] Dias, C. P. (2000) - *Análise Conjunta Pesada de Regressões* Master's Thesis. Évora University.
- [6] Digby, P. G. N. (1979) - *Modified joint regression analysis for incomplete variety \times environment data* - J. Agric. Sci., Camb., 93 : 81 - 86.
- [7] Eberhart, S. A. and Russel, W. A. (1966) - *Stability parameters for comparing varieties* - Crop. Sci, 6 : 36 - 40.
- [8] Ferreira, T.; Mexia, J. T.; Baeta, J.; Maças, B.; Gusmão, L.; Bagulho, F. e Coutinho, J. (1990) - *Comportamento relativo das cultivares de triticales produzidas na E.N.M.P., no período 1987/88 e 1988/89: II. Aferido pela análise de regressão conjunta*. Em "V Reunião Portuguesa sobre Triticale- Elvas, Maio 1990 - Melhoramento, vol. 32.
- [9] Finlay, K. W. and Wilkison, G. N. (1963) - *The analysis of adaptation in a plant-breeding programme*. - Aust. J.Aric. Res., 14 : 742 - 754.
- [10] Fisher, R. A. (1918) - *The correlation between relatives on the supposition of Mendelian inheritance* - Transactions of the Royal Society of Edinburgh, 42: 399 - 433.

- [11] Fonseca, M.; Mexia, J.T. and Zmyslony, R. (2003a) - *Estimators and Tests for variance Components in Cross Nested Orthogonal Models*. Discussion in Mathematics, Probability and Statistics 23: 173 - 201.
- [12] Gusmão, L. (1985) - *An adequate design for regression analysis of yield trials*. Theor. Appl. Genet., 71: 314 - 319.
- [13] Gusmão, L. (1986a) - *Inadequacy of blocking in cultivar yield trials*. Theor. Appl. Genet., 72: 98 - 104.
- [14] Gusmão, L. (1986b) - *A interação genótipo x ambiente e a comparação de cultivares de cereais*. Tese de Doutoramento. Instituto Superior de Agronomia. Universidade Técnica de Lisboa, Lisboa.
- [15] Gusmão, L. (1990) - *Avaliação de cultivares (uma perspectiva experimental diferente)*. Revista de Ciências Agrárias - n°2, vol. 13: 11 - 18.
- [16] Gusmão, L.; Baeta, J. e Mexia, J.T. (1991a) - *Avaliação da aptidão agronómica relativa em cultivares de cereais*. In "I Jornadas de Estatística e Aplicações". Universidade do Minho - Braga, 22 a 24 de Abril.
- [17] Gusmão, L.; Rebelo, D. C.; Miranda, M. J. e Silva, A. F. (1991b) - *Year interaction on interplant competition in white Lupin (*Lupinus Albus L.*)*. Jn: Martins, J.M.N. and Costa, M.L.(eds) Advances in Lupin Research Rd. ISA, Lisboa, Pág. 409 - 412.
- [18] Gusmão, L.; Mexia, J.T. e Baeta, J. (1992a) - *Trimmed Joint Regression: a new approach to the Joint Regression Analysis for cultivar relative performance evaluation*. Theoretical and Applied Genetics, 84 : 735 - 738.
- [19] Gusmão, L.; Baeta, J.; Mexia J.T.; Bagulho, F.; Maçãs, B. e Coutinho, J. (1992b) - *Statistical analysis of randomized complete block designs for Barley cultivar yield trials*. Rachis, 11 : 14 - 18.
- [20] Gusmão, L.; Rebelo, D. C.; Miranda, M.J.; Baeta, J. e Vaz, F. (1993a) - *Assessment of specific genotype × environment interactions in *Lupinus luteus L.** Brotéria Genética, Lisboa, 14: 141 - 144.
- [21] Gusmão, L.; Rebelo, D. C. e Mexia J.T. (1993b) - *Increasing the precision of relative yield behaviour evaluation of *Lupinus Luteus L.* cultivars, in the presence of specific interactions*. ABST. - VII International Lupin Conference, Évora, 18 a 23 de Outubro, Theme:3 , n° 32.

- [22] Gusmão, L.; Fernandes, D.; Maças, B.; Mexia J.T.; Gusmão, M. R. (1995) - *Análise pela "RECOR" dos resultados de produção numa rede regional de ensaios de aveias (Avena Sativa L.)* da E.N.M.P. de 1990/91 a 1991/92. *Revista de Ciências Agrárias*, vol. 18, 2: 25 - 33.
- [23] Ito, K. (1980) - *Robustness of Anova and Macanova test procedures*. P.R. Krishnaiah (ed) - *Handbook of Statistics - Vol. I - North Holland*
- [24] John, J. A. and Williams, E. R. (1995) - *Cyclic and computer generated designs*. Chapman and Hall, London, 2nd ed..
- [25] Judge, George G.; Hill, R. Carter; Griffiths, William E.; Lütkepohl, Helmut L. and Lee, Tsoung-Chao, (1988) - *Introduction to the theory and practise of econometrics*. John Wiley and Sons, Inc. 2nd edition.
- [26] Kageyama, S. (1984) - *Connected designs with the minimum number of experiments units*. *Linear Statistical Inference, Lecture Notes in Statistics* n. 35, Springer Verlag, New York : 99 - 117.
- [27] Lin, C. S., Binns, M. R. and Lefkovitch, L. P. (1986) - *Stability analysis: Where Do We Stand?* *Crop. Sci.*, 26: 894 - 900.
- [28] Maças, B.; Coutinho, J.; Bagulho, F. Mexia, J. T.; Baeta, J.; Gusmão, L. e Ferreira, T. (1990) - *Comportamento relativo das cultivares de triticales produzidas na E.N.M.P., no período 1987/88 e 1988/89: I. Aferido pela análise convencional*. Em "V Reunião Portuguesa sobre Triticales- Elvas, Maio 1990 - Melhoramento, vol. 32.
- [29] Mexia, J. T. (1987) - *Multi-treatment Regression Designs* *Trabalhos de Investigação*, n° 1, 1987 - Departamento de Matemática, FCT-UNL.
- [30] Mexia, J. T. (1989) - *Controlled Heterocedasticity, Quocient Vector Spaces and F tests for hypothesis on Mean Vectors* - *Trabalhos de Investigação*, n° 2, 1989 - Departamento de Matemática, FCT-UNL.
- [31] Mexia, J. T. (1990) - *Best Linear unbiased estimates, duality of F tests and the Scheffé multiple comparison method in the presence of controlled heterocedasticity*. *Comp. Stat and Data Analysis - Vol.10, n° 3*.
- [32] Mexia, J. T.; Gusmão, L. e Baeta, J. (1991) - *Analysis of cultivar yield trials designed in Randomize Complete Blocks*. - *Revista de Ciências Agrárias*, 14: 59 - 63.
- [33] Mexia, J. T., (1992) - *Asymptotics Chi-Squared Tests, Designs and Log-Linear models*. *Trabalhos de Investigação*. N°1, 1992.

- [34] Mexia, J. T. (1995) - *Introdução à Inferência Estatística Linear*. Edições Lusófonas, Lisboa.
- [35] Mexia, J. T.; Amaro, A. P., Gusmão, L. e Baeta J. (1997) - *Upper Contour of a Joint Regression Analysis*. J. Genet. and Breed., 51 : 253 - 255.
- [36] Mexia, J. T.; Pereira, D. G. e Baeta, J. (1999) - *L₂ environmental indexes*. Biometrical Letters, 36: 137 - 143.
- [37] Mexia, J. T. e Pinto, I. (2003) - *Joint Regression Analysis and Connectedness - Application to Grapevines*- Colloquium Biometryczne, 33: 235 - 248.
- [38] Mood, A. M.; Graybill, F. A. and Boes, D. C. (1974) - *Introduction to the Theory of Statistics*. McGraw-Hill, Singapura, 3rd edition.
- [39] Mooers, C. A. (1921) - *The agronomic placement of varieties*. - J. Amer. Soc. Aron., 13 : 337 - 352.
- [40] Ng, M. P. and Grunwald, G. K. (1997) - *Nonlinear regression analysis of the joint-regression model*. - Biometrics, 53 : 1366 - 1372.
- [41] Patterson, H.D. and Williams, E.R. (1976) - *A new class of resolvable incomplete block designs*. Biometrika, 63: 83 - 92.
- [42] Pereira, D.G. (2004) - *Análise Conjunta Pesada de Regressões em Redes de Ensaios*. - Tese de Doutorado Universidade de Évora.
- [43] Raghavarao, D. (1971) - *Construction and combinatorial Problems in Design of Experiments*. Wiley, New York. Reprinted (1988) by Dover with some addendum.
- [44] Scheffé, H. (1959) - *The Analysis of Variance*. John Willey and Sons - Nova York.
- [45] Seber, G.A.F. (1980) - *The Linear Hypothesis : A General theory*, 2nd ed., Charles Griffin and Co - Londres.
- [46] Seber, G.A.F. and Lee, J. A. (2003) - *Linear Regression Analysis*, 2nd ed., John Willey and Sons - New York.
- [47] Severini, T. A. (2000) - *Likelihood methods in statistics*. Oxford University Press, New York.
- [48] Silvey, S. D. (1975) - *Statistical Inference*- Chapman and Hall, Cambridge.
- [49] Siegel, S. and Castellan, J. (1988) - *Nonparametric statistics for the behavioral science*.- 2nd ed., Mc Graw-Hill, New York.

-
- [50] Westcott, B. (1986) - *Some methods of analysing genotype-environment interaction*
Heredity, 56 : 243 - 253.
- [51] Yates, F. and Cochran, W. G. (1938) - *The analysis of groups experiments*. J. Agric.
Sci., Cambridge, 28 : 556 - 580.