

UNIVERSIDADE ABERTA

ESTABILIDADE EM ANÁLISE CONJUNTA
DE REGRESSÕES E CONDUÇÃO DINÂMICA
DE PLANOS DE MELHORAMENTO

AMÍLCAR MANUEL DO ROSÁRIO OLIVEIRA

Dissertação apresentada para
obtenção do Grau de Doutor em
Matemática - Especialidade em
Modelação Estatística pela Uni-
versidade Aberta.

"Esta tese não inclui as críticas e sugestões feitas pelo júri"

Lisboa

2007

UNIVERSIDADE ABERTA

ESTABILIDADE EM ANÁLISE CONJUNTA
DE REGRESSÕES E CONDUÇÃO DINÂMICA
DE PLANOS DE MELHORAMENTO

AMÍLCAR MANUEL DO ROSÁRIO OLIVEIRA

Orientador: Professor Doutor João Tiago Praça Nunes Mexia

Co-orientador: Professor Doutor Alexandre Gomes Cerveira

Lisboa

2007

*Na criação do Cosmos, os pensamentos de Deus
são pensamentos matemáticos*

Santo Agostinho

*Dedico este trabalho à minha esposa
e aos meus filhos*

Agradecimentos

Ao Professor Doutor João Tiago Mexia, meu orientador, agradeço todo o apoio e incentivo que me transmitiu em todos os momentos. Agradeço-lhe ainda pelos ensinamentos, pela dedicação, pela disponibilidade, para trabalharmos na elaboração desta dissertação.

Ao Professor Doutor Alexandre Cerveira, agradeço todo o apoio e interesse que manifestou pelo trabalho desenvolvido.

À minha esposa, pela força que me transmitiu, pelo apoio incondicional e acompanhamento em todas as circunstâncias.

Aos meus filhos Joana e Guilherme, agradeço pela compreensão que tiveram pela minha ausência em muitos momentos e pelo interesse demonstrado pelo meu trabalho.

Ao Professor Doutor Stanislaw Mejza, pelas sugestões e críticas construtivas efectuadas.

Ao Professor Doutor Tadeusz Calinski, pelos seus comentários e sugestões.

Ao Professor Doutor Christopher Minder, pelo interesse manifestado e pelas sugestões dadas.

Às colegas Rosário Ramos e Lucinda Fernandes e ao colega Mário Edmundo agradeço pelo trabalho adicional que tiveram enquanto estive ausente em dispensa de serviço docente.

À Iola Pinto, à Dulce Pereira, ao Dário Ferreira, à Sandra Ferreira, ao Luis Ramos, ao Rafael Sasportes e a todos os colegas e amigos que de alguma forma me deram o seu apoio.

Ao Departamento de Cereais da Estação Nacional de Melhoramento de Plantas, em particular ao Senhor Engenheiro Bemvindo Maçãs e ao Senhor Engenheiro João Coco, pela disponibilização dos dados utilizados neste trabalho.

À Universidade Aberta, instituição que me concedeu a dispensa de serviço docente tendo em vista a realização desta dissertação.

Resumo

A Análise Conjunta de regressões, ACR, tem sido uma técnica bastante utilizada no domínio da avaliação de cultivares.

O principal objectivo desta dissertação é a extensão da aplicação da técnica ACR, na medida em que passamos duma análise usual, em que se tem em atenção apenas as produções de cada cultivar, para o caso em que se considera o valor económico, traduzido pelo peso específico.

A aplicação da técnica consistia até agora, no ajustamento de regressões lineares da produção de cada cultivar no índice ambiental.

Aprofundaremos a sua aplicação a planos de melhoramento, logo a casos em que se tem um número alargado de anos de ensaios efectuados.

Abordaremos a problemática associada à convergência do algoritmo zig-zag, apresentando uma validação de ajustamentos efectuada, através dum modelo linear para os resíduos.

Daremos ênfase ao conceito de quase normalidade multiplicativa, através de um conjunto de simulações que nos permitirão tirar conclusões importantes tendo em vista a aplicação prática.

Abstract

Joint Regression Analysis, JRA, has been a widely used technique on the evaluation of cultivars.

The aim of this dissertation is to extend the application of JRA, as we go from a regular analysis in which is paid attention only to the production of each cultivar, to the case in which is considered the economic value, given by the specific weight. The application of this technique so far consisted in the adjustment of linear regressions of the production of each cultivar on the environmental index. We will deepen its application to improvement plans, so to cases in which we have a wide number of years of trials done.

We will discuss the problems related to the convergence of the zig-zag algorithm, presenting a validation of adjustments through a linear model for the residues.

We will emphasise the concept of the multiplicative approximate normality through a set of simulations which will allow us to achieve important conclusions concerning the practical application.

Lista de símbolos e siglas

- ACR Análise conjunta de regressão
- β coeficiente angular da regressão linear
- $Var(U)$ variância da variável U
- $Var(V)$ variância da variável V
- $E(U)$ valor esperado da variável U
- $E(V)$ valor esperado da variável V
- μ_ℓ valor médio da variável X_ℓ
- σ_ℓ desvio padrão da variável X_ℓ
- L_2 Índices ambientais L_2
- J número de cultivares estudados
- $p_{i,j}; i = 1, \dots, b; j = 1, \dots, J$ peso atribuído ao j -ésimo cultivar no i -ésimo bloco
- $p_i; i = 1, \dots, b$ peso atribuído a um cultivar que esteja presente no bloco i
- b número de blocos

- $y_{i,j}; i = 1, \dots, b; j = 1, \dots, J$ produção do j -ésimo cultivar no i -ésimo bloco
- α_j ordenada na origem da recta de regressão a ajustar do j -ésimo cultivar
- β_j coeficiente angular da recta de regressão a ajustar do j -ésimo cultivar
- x_i índice ambiental do i -ésimo bloco
- α_J vector das ordenadas na origem para as regressões
- β_J vector dos coeficientes angulares para as regressões
- \mathbf{x}_b vector dos índices ambientais
- $S_1(\alpha_J, \beta_J)$ função objectivo a minimizar na primeira iteração do algoritmo zig-zag
- $\mathbb{R}_b \times \mathbb{R}_J \times \mathbb{R}_J$ espaço obtido pelo produto cartesiano
- x_i^0 valor inicial do índice ambiental do i -ésimo bloco
- $R(\mathbf{1}_b)$ espaço imagem dos vectores coluna com todas as componentes iguais a 1
- $grad(S)$ gradiente da função S
- $Hess(S)$ matriz hessiana de S

- $(\tilde{\alpha}_j, \tilde{\beta}_j)$ coeficientes de regressão ajustados referentes ao j-ésimo cultivar
- \tilde{x}_i índice ambiental ajustado no i-ésimo bloco
- $r_{i,j}$ resíduo associado ao j-ésimo cultivar no i-ésimo bloco
- L número de sub-redes conectadas
- A_ℓ conjunto dos índices dos blocos para a l-ésima sub-rede conectada, $\ell = 1, \dots, L$
- $(\tilde{\alpha}_j^1, \tilde{\beta}_j^1)$ estimadores dos mínimos quadrados para α_j^1, β_j^1 obtidos na primeira iteração do algoritmo zig-zag
- $S_1(\mathbf{x}_b)$ função a minimizar em ordem a \mathbf{x}_b na primeira iteração do algoritmo zig-zag
- $Int(a)$ parte inteira de a
- $A \otimes B$ produto de Kronecker de A por B
- \mathbf{x}'_b transposta de \mathbf{x}_b
- \mathbf{Y}' transposta de \mathbf{Y}
- $L(\cdot)$ função de máxima verosimilhança
- D_i conjunto dos índices dos cultivares presentes no i-ésimo bloco, $i = 1, \dots, b$

- N número de observações
- $\ln(\cdot)$ logaritmo da função de verosimilhança
- $\hat{\alpha}^2$ estimador de máxima verosimilhança para σ^2
- $(\hat{\alpha}_j, \hat{\beta}_j)$ estimadores de máxima verosimilhança
- $\mathbf{1}_J$ vector com J componentes iguais a 1
- $\tilde{\theta}_{j,j'}$ ponto de intersecção entre as rectas de regressão ajustadas com índices j e j'
- $H_0^{j,j'}(\theta_0)$ hipótese nula relativa aos $j - \text{imo}$ e $j' - \text{imo}$ cultivares e à função $U(\theta_0)$
- $t_{j,j'}(\theta_0)$ valor da estatística do teste t de Student para a comparação dos cultivares j e j'
- $f_{1-\alpha, g_1, g}$ valor crítico correspondente à distribuição F de Snedcor central, com g_1 e g graus de liberdade, para o nível de significância α
- $N(\underline{\eta}_m, \sigma^2 K)$ distribuição normal, com valor médio $\underline{\eta}_m$ e matriz de covariância $\sigma^2 K$
- χ_g^2 distribuição qui-quadrado central, com g graus de liberdade
- $D(r_1, \dots, r_m)$ matriz diagonal do tipo sxs , cujos elementos da diagonal principal correspondem a r_1, \dots, r_m
- S_{a_ℓ} soma dos quadrados dos desvios para a média dos índices ambientais ajustados no ano com índice a_ℓ

Conteúdo

1	Introdução	1
2	Quase-Normalidade Multiplicativa	2
2.1	Considerações prévias	2
2.2	Convergência quase certa	3
2.3	Simulações	5
3	Análise Conjunta de Regressões	11
3.1	Enquadramento histórico da ACR	11
3.2	Considerações prévias	13
3.3	Variáveis não observáveis	14
3.4	Caso geral	14
3.4.1	Função Objectivo	15
3.4.2	Redes Conectadas	22
3.4.3	Algoritmo zig-zag	23
3.4.4	Análise dos Resíduos	26
3.5	Caso completo	32
3.5.1	Dupla minimização	32
3.5.2	Estimadores de máxima verosimilhança	35
3.5.3	Testes de razão de verosimilhanças	36
3.6	Redes Fraccionalmente Completas	42
3.7	Método do Contorno Superior	43
3.7.1	Geometria do contorno superior	44
3.7.2	Comparação e selecção de cultivares	48

Conteúdo	x
3.8 Estabilidade dos Cultivares	52
4 Análise Conjunta de Regressões Lineares Completadas	54
4.1 Cultivares com Resposta Óptima	54
4.2 Método do Contorno Superior	55
4.2.1 Reparametrização e intersecção de rectas	55
4.2.2 Reparametrização e intersecção de várias rectas	58
4.3 Comparações múltiplas para os coeficientes corrigidos	59
5 Plano de Melhoramento de Aveia	61
5.1 Considerações Prévias	61
5.2 Análise do contorno superior	70
5.3 Contorno superior e selecção	71
5.4 Classificação e selecção dos cultivares com coeficientes angulares corrigidos	73
5.5 Notas críticas	75
6 Desenvolvimentos futuros	76
A Introdução aos α-Designs	77
B Gráficos com rectas ajustadas	79
C Programas de Simulação	90
C.1 Grupo 1	90
C.2 Grupo 2	91
C.3 Grupo 3	92
Bibliografia	94

Lista de Tabelas

2.1	Número de rejeições para $(\alpha = 1\%), \sigma_1 = \sigma_2 = 1$	5
2.2	Número de rejeições para $(\alpha = 1\%), \sigma_1 = \sigma_2 = 10$	6
2.3	Número de rejeições para $(\alpha = 1\%), \sigma_1 = \sigma_2 = 100$	6
3.1	Locais e anos dos ensaios	28
3.2	Coefficientes de regressão ajustados	29
3.3	Índices ambientais	30
3.4	Coefficientes ajustados	31
3.5	Somas de quadrados dos resíduos e coeficientes R^2	31
5.1	Lista dos cultivares ensaiados	62
5.2	Locais e anos onde se realizaram os ensaios	63
5.3	Presença/ausência dos cultivares nos vários anos	64
5.4	Médias dos índices ambientais por bloco	65
5.5	Soma dos quadrados dos Resíduos para a média dos índices ambientais	65
5.6	Coefficientes angulares ajustados	66
5.7	Coefficientes angulares ajustados (cont.)	67
5.8	Soma dos quadrados dos Resíduos para a média dos índices ambientais obtidos	68
5.9	Resultados obtidos através dos testes F	69
5.10	Composição dos contornos superiores ano a ano	70
5.11	Resultados obtidos no ajustamento	71
5.12	Cultivares dominantes e dominados: resultados da aplicação do teste t	72

5.13 Resultados dos métodos de comparação múltipla de Scheffé e de Bonferroni ao nível de 5% - Anos 1994,1995,1996,1997	73
5.14 Comparações múltiplas com coeficientes corrigidos	74
5.15 Resultados do método aproximado de comparação múltipla de Scheffé ao nível de 5% - Anos 1994,1995,1996,1997	74

Lista de Figuras

2.1	Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 1$; $\mu_1 = 1, 2, \dots, 100$; $\mu_2 = 1, 2, \dots, 20$	7
2.2	Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 10$; $\mu_1 = 1, 2, \dots, 100$; $\mu_2 = 1, 2, \dots, 20$	7
2.3	Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 100$; $\mu_1 = 1, 2, \dots, 100$; $\mu_2 = 1, 2, \dots, 20$	8
2.4	Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 1$; $\mu_\ell = 1 + 100.i$; $i = 0, 1, 2, \dots, 100$; $\ell = 1, 2$	9
2.5	Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 10$; $\mu_\ell = 1 + 100.i$; $i = 0, 1, 2, \dots, 100$; $\ell = 1, 2$	9
2.6	Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 100$; $\mu_\ell = 1 + 100.i$; $i = 0, 1, 2, \dots, 100$; $\ell = 1, 2$	10
3.1	Rectas paralelas	44
3.2	Intersecção das rectas ajustadas de índices j e j'	45
3.3	Traçado do contorno superior para $\bar{\theta}_1 \leq \theta_{min}$	46
3.4	Traçado do contorno superior para $\theta_{min} < \bar{\theta}_1 \leq \theta_{max}$	46
3.5	Traçado do contorno superior para $\bar{\theta}_1 \geq \theta_{max}$	47
4.1	Gráfico da função $x(c)$ com $\theta_1 < \theta_2$	56
4.2	Gráfico da função $x(c)$ com $\theta_1 > \theta_2$	56
4.3	Representação gráfica da intersecção das rectas quando $\theta_1 < \theta_2$	56
4.4	Intersecção de várias rectas	58
B.1	Análise conjunta de regressões - 1984	79
B.2	Análise conjunta de regressões - 1985	80

B.3	Análise conjunta de regressões - 1986	80
B.4	Análise conjunta de regressões - 1987	81
B.5	Análise conjunta de regressões - 1988	81
B.6	Análise conjunta de regressões - 1989	82
B.7	Análise conjunta de regressões - 1990	82
B.8	Análise conjunta de regressões - 1991	83
B.9	Análise conjunta de regressões - 1992	83
B.10	Análise conjunta de regressões - 1993	84
B.11	Análise conjunta de regressões - 1994	84
B.12	Análise conjunta de regressões - 1995	85
B.13	Análise conjunta de regressões - 1996	85
B.14	Análise conjunta de regressões - 1997	86
B.15	Análise conjunta de regressões - 1998	86
B.16	Análise conjunta de regressões - 1999	87
B.17	Análise conjunta de regressões - 2000	87
B.18	Análise conjunta de regressões - 2002	88
B.19	Análise conjunta de regressões - 2003	88
B.20	Análise conjunta de regressões - 2004	89

Capítulo 1

Introdução

A análise conjunta de regressões, ACR, é uma técnica que tem sido utilizada na análise conjunta de ensaios de comparação de cultivares.

Esta técnica foi inicialmente utilizada para ensaios com os mesmos cultivares sendo depois, ver Pinto(2006), extendida à condução de planos de melhoramento. Nestes planos haverá entrada e saída de cultivares.

A variável base das aplicações da ACR tem sido o rendimento. Vamos agora ver como é possível incorporar na mesma informação adicional que se tenha, o valor das produções dos diferentes cultivares. Tipicamente esta informação só estará disponível para uma fracção pequena de resultados.

Ora na ACR ajusta-se por cultivar uma regressão linear da produção numa variável, o índice ambiental, que mede a capacidade produtiva. Assim o coeficiente angular β medirá a taxa de aumento da produção com a melhoria das capacidades. Se se tiver uma medida V do valor da produção, βV medirá a taxa de aumento da produção com a capacidade produtiva. Dado que actualmente se valorizam cultivares com boa capacidade de aproveitar condições de alta fertilidade é natural utilizar os valores de βV para comparar os cultivares.

A fim de poder realizar inferência começamos por mostrar, no capítulo 2, que o produto de variáveis normais independentes uma das quais com baixo coeficiente de variação é aproximadamente normal.

Em seguida no capítulo 3 apresentamos a ACR clássica mostrando como a aplicar a um plano de melhoramento.

Segue-se o capítulo 4 onde se mostra como aplicar a ACR completada considerando-se a informação adicional.

Finalmente no capítulo 5 aplica-se a técnica desenvolvida ao plano português de melhoramento da aveia (1984-2004).

Capítulo 2

Quase-Normalidade Multiplicativa

2.1 Considerações prévias

Dadas duas variáveis normais independentes X_1 e X_2 com valores médios μ_1 e μ_2 e desvios padrões σ_1 e σ_2 respectivamente, escrevendo-se $X_\ell \sim N(\mu_\ell, \sigma_\ell)$; $\ell = 1, 2$, vamos mostrar que quando o coeficiente de variação

$$c_1 = \frac{\sigma_1}{\mu_1}$$

é suficientemente pequeno $X_1 X_2$ é aproximadamente normal com valor médio $\mu_1 \mu_2$ e desvio padrão $\mu_1 \sigma_2$, escrevendo-se

$$X_1 X_2 \sim N(\mu_1 \mu_2, \mu_1 \sigma_2)$$

Com efeito tem-se $X_1 = \dot{X} + \mu_1$ com $\dot{X} \sim N(0, \sigma_1)$, pelo que

$$X_1 X_2 = \dot{X} X_2 + \mu_1 X_2$$

tendo-se $\mu_1 X_2 \sim N(\mu_1 \mu_2, \mu_1 \sigma_2)$.

Ora na secção seguinte mostramos que, representando por \xrightarrow{qc} a convergência quase certa,

$$\frac{X_1 X_2}{\mu_1 X_2} \xrightarrow[\mu_1 \rightarrow +\infty]{qc} 1$$

em condições extremamente gerais.

Em seguida na alínea 2.3 apresentaremos os resultados de simulações que mostram que, no caso de produtos de variáveis normais independentes, não é preciso exigir valores muito baixos para c_1 para que o produto possa ser tratado como normal.

2.2 Convergência quase certa

Começemos por estabelecer

Proposição 2.1 *Consideremos $U = (X_1 + b_n)X_2$ e $V = b_n X_2$, com X_1 e X_2 variáveis aleatórias $X_1 \sim (\mu_1, \sigma_1)$ independente de $X_2 \sim (\mu_2, \sigma_2)$ com b_n constante. Tem-se*

$$\frac{Var(U)}{Var(V)} \xrightarrow[b_n \rightarrow \infty]{} 1 \quad (2.2.1)$$

Dem: Facilmente se obtêm, os valores médios $E(U) = E[(X_1 + b_n)X_2] = (\mu_1 + b_n)\mu_2$ e $E(V) = E(b_n X_2) = b_n E(X_2) = b_n \mu_2$.

Por outro lado temos as variâncias

$$\begin{aligned} Var(U) &= (b_n^2 + \sigma_1^2)(\mu_2^2 + \sigma_2^2) - [(\mu_1 + b_n)\mu_2]^2 \\ &= b_n^2 \mu_2^2 + b_n^2 \sigma_2^2 + \sigma_1^2 - (\mu_1 + b_n)^2 \mu_2^2 \\ &= b_n^2 \sigma_2^2 + \sigma_1^2 \mu_2^2 + \sigma_1^2 \sigma_2^2 - \mu_1^2 \mu_2^2 - 2b_n \mu_1 \mu_2^2 \end{aligned}$$

e

$$Var(V) = b_n^2 Var(X_2) = b_n^2 \sigma_2^2$$

então virá

$$\frac{Var(U)}{Var(V)} = 1 + \frac{\sigma_1^2 \mu_2^2}{b_n^2 \sigma_2^2} + \frac{\sigma_1^2}{b_n^2} - \frac{\mu_1^2 \mu_2^2}{b_n^2 \sigma_2^2} - \frac{2\mu_1 \mu_2^2}{b_n \sigma_2^2}$$

pelo que

$$\frac{\text{Var}(U)}{\text{Var}(V)} \xrightarrow{b_n \rightarrow \infty} 1 \quad (2.2.2)$$

■

Representemos agora a convergência quase certa por \xrightarrow{qc} e estabeleçamos a

Proposição 2.2 *Se $b_n \rightarrow +\infty$, qualquer que seja a variável aleatória Z tem-se*

$$\frac{Z}{b_n} \xrightarrow{b_n \rightarrow +\infty} 0$$

Dem: Qualquer que seja $k > 0$ existe \bar{n} tal que, para $n > \bar{n}$, $b_n > k$ pelo que, com $\epsilon > 0$, $-k\epsilon \leq Z_n \leq k\epsilon$ implicará $\bigcap_{n=\bar{n}}^{+\infty} (-b_n\epsilon \leq Z_n \leq b_n\epsilon)$, tendo-se, com F_Z a distribuição de Z

$$\text{pr}\left(\bigcap_{n=\bar{n}}^{+\infty} \left|\frac{Z_n}{b_n}\right| \leq \epsilon\right) = \text{pr}\left(\bigcap_{n=\bar{n}}^{+\infty} (-b_n\epsilon \leq Z_n \leq b_n\epsilon)\right) \geq \text{pr}(-k\epsilon \leq Z_n \leq k\epsilon) = F(k\epsilon) - F(-k\epsilon)$$

Assim qualquer que seja k , ter-se-á

$$\text{pr}\left(\bigcup_{m=1}^{\infty} \bigcap_{m=n}^{\infty} \left|\frac{Z_n}{b_n}\right| \leq \epsilon\right) \geq F(k\epsilon) - F(-k\epsilon)$$

vindo, qualquer que seja $\epsilon > 0$

$$\text{pr}\left(\bigcup_{m=1}^{\infty} \bigcap_{m=n}^{\infty} \left|\frac{Z_n}{b_n}\right| \leq \epsilon\right) = 1$$

Para completar a demonstração basta observar que, ver Loève(1961),

$\bigcup_{m=1}^{\infty} \bigcap_{m=n}^{\infty} \left(\left|\frac{Z_n}{b_n}\right| \leq \epsilon\right)$ é o limite inferior da sucessão $\{A_n(\epsilon)\}$ de acontecimentos cujo termo geral verifica-se se e só se $\left|\frac{Z_n}{b_n}\right| \leq \epsilon$. ■

Corolário 2.1 *Dado um par qualquer X_1, X_2 de variáveis aleatórias normais independentes tem-se, caso $b_n \rightarrow +\infty$,*

$$\frac{(X_1 + b_n)X_2}{b_n X_2} \xrightarrow{b_n \rightarrow +\infty} 1$$

Dem: Basta observar que

$$\frac{(X_1 + b_n)X_2}{b_n X_2} = 1 + \frac{X_1}{b_n}$$

Além disso as simulações efectuadas mostram que se $X_\ell \sim N(\mu_\ell, \sigma_\ell)$; $\ell = 1, 2$, e o coeficiente de variação for suficientemente pequeno então $X_1 X_2 \sim N(\mu_1 \mu_2, \mu_1 \sigma_2)$.

■

2.3 Simulações

Apresentam-se nesta secção os resultados dum conjunto de simulações realizadas utilizando a aplicação R. Para cada combinação de parâmetros foram geradas 1000 amostras constituídas por 100 pares de valores para X_1 e X_2 , dos quais resultam 100 valores para $U = X_1 X_2$. A normalidade dessas amostras foi testada usando o teste Kolmogorov-Smirnov correspondente. Na geração das amostras assumimos que $X_\ell \sim N(\mu_\ell, \sigma_\ell)$, $\ell = 1, 2$ tendo sido consideradas as propriedades bem conhecidas da distribuição normal.

No primeiro conjunto de simulações tomámos $\mu_\ell \in \{1, 2, 5, 7, 10, 15, 20, 25, 30\}$ e $\sigma_\ell \in \{1, 10, 100\}$, $\ell = 1, 2$. Os números de rejeições ao nível de significância de 1% são apresentados nas tabelas 2.1, 2.2 e 2.3. Na primeira e segunda tabelas verificamos que quando μ_1 predomina sobre μ_2 o número de rejeições cai claramente.

Tal queda não se verifica no caso apresentado na Tabela 2.3. Este comportamento é concordante com o que seria de esperar, uma vez que os coeficientes de variação são muito maiores no caso da tabela 2.3.

Observe-se que o predomínio de μ_1 sobre μ_2 corresponderá a ter-se coeficiente de variação menor para X_1 do que para X_2 .

Tabela 2.1: Número de rejeições para $(\alpha = 1\%), \sigma_1 = \sigma_2 = 1$

μ_1	μ_2								
	1	2	5	7	10	15	20	25	30
1	756	999	1000	1000	1000	1000	1000	1000	1000
2	123	463	1000	1000	1000	1000	1000	1000	1000
5	12	22	218	714	994	1000	1000	1000	1000
7	8	11	64	216	761	998	1000	1000	1000
10	8	10	17	65	214	807	996	1000	1000
15	5	8	12	13	44	211	621	928	992
20	4	7	8	12	20	69	189	500	800
25	7	16	10	10	11	26	75	197	433
30	9	5	6	9	15	20	40	100	185

Tabela 2.2: Número de rejeições para $(\alpha = 1\%), \sigma_1 = \sigma_2 = 10$

μ_1	μ_2								
	1	2	5	7	10	15	20	25	30
1	1000	1000	1000	1000	1000	1000	1000	1000	1000
2	1000	1000	1000	1000	1000	1000	1000	1000	1000
5	235	279	696	946	1000	1000	1000	1000	1000
7	54	96	472	740	956	999	1000	1000	1000
10	66	100	320	501	759	966	996	1000	1000
15	61	50	136	204	335	567	864	960	996
20	38	35	62	80	154	286	460	720	882
25	21	19	23	41	62	132	233	377	609
30	9	7	16	15	33	57	107	184	343

Tabela 2.3: Número de rejeições para $(\alpha = 1\%), \sigma_1 = \sigma_2 = 100$

μ_1	μ_2								
	1	2	5	7	10	15	20	25	30
1	1000	1000	1000	1000	1000	1000	1000	1000	1000
2	1000	1000	1000	1000	1000	1000	1000	1000	1000
5	1000	1000	1000	1000	1000	1000	1000	1000	1000
7	1000	1000	1000	1000	1000	1000	1000	1000	1000
10	1000	1000	1000	1000	1000	1000	1000	1000	1000
15	1000	1000	1000	1000	1000	1000	1000	1000	1000
20	1000	999	1000	999	1000	999	999	1000	1000
25	996	997	991	996	992	988	993	996	997
30	916	924	934	930	926	937	944	949	967

Em seguida realizou-se uma segunda série de simulações, tomando como anteriormente $\sigma_1 = \sigma_2 \in \{1, 10, 100\}$. Agora $\mu_1[\mu_2]$ toma os valores inteiros de 1 a 100[1 a 20]. Neste caso optámos por uma representação gráfica dos resultados que se apresentam nas figuras 2.1, 2.2 e 2.3.

Figura 2.1: Número de rejeições para $(\alpha = 1\%); \sigma_1 = \sigma_2 = 1; \mu_1 = 1, 2, \dots, 100; \mu_2 = 1, 2, \dots, 20$

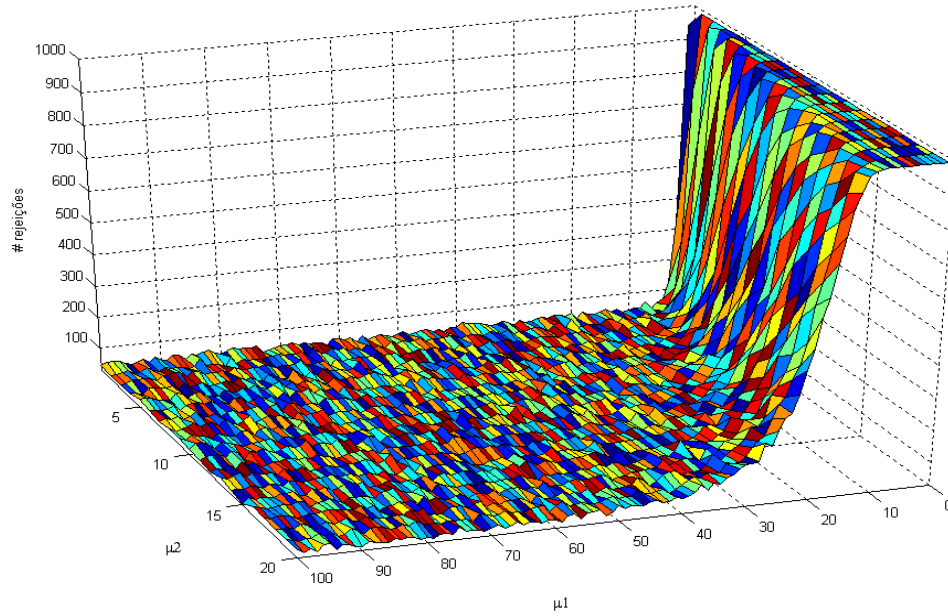


Figura 2.2: Número de rejeições para $(\alpha = 1\%); \sigma_1 = \sigma_2 = 10; \mu_1 = 1, 2, \dots, 100; \mu_2 = 1, 2, \dots, 20$

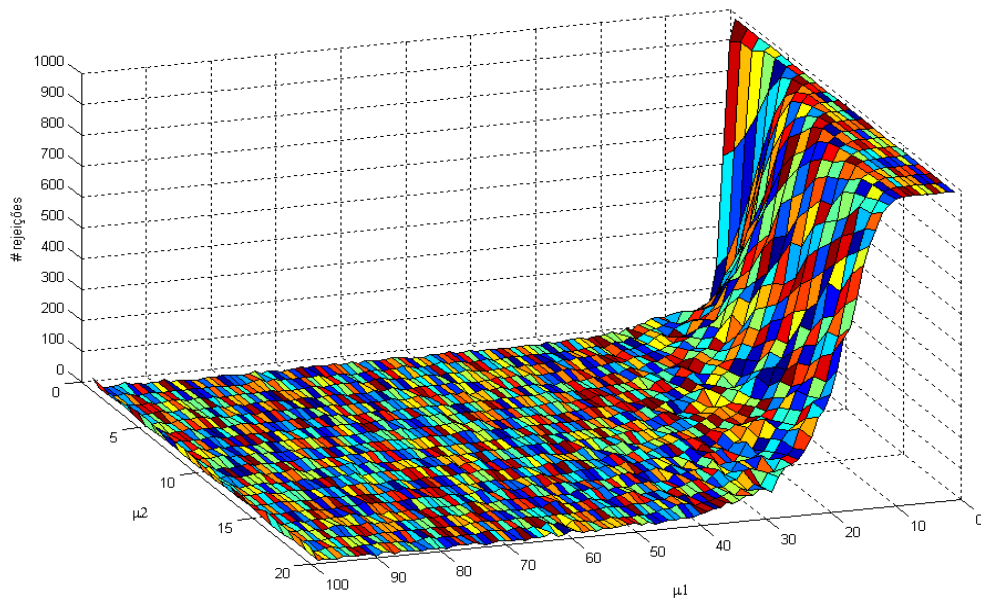
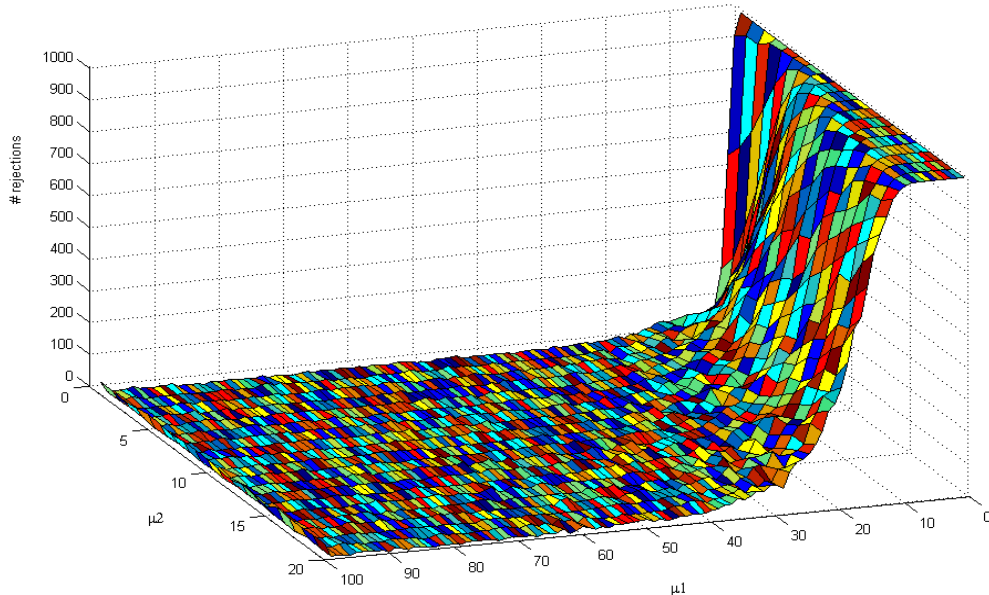


Figura 2.3: Número de rejeições para $(\alpha = 1\%); \sigma_1 = \sigma_2 = 100; \mu_1 = 1, 2, \dots, 100; \mu_2 = 1, 2, \dots, 20$



A queda abrupta no número de rejeições é mais uma vez bem visível. A geometria associada a esta queda revela-se interessante. Quando se consideram valores crescentes de $\theta = \frac{\mu_1}{\mu_2}$ tem-se um primeiro conjunto para o qual existem muitas rejeições, após isso tem-se a queda. A projecção desta queda na horizontal situa-se aproximadamente ao longo da linha recta $\mu_1 = \mu_2$. Para completar o estudo da aproximação à normalidade tomou-se um terceiro conjunto de simulações, considerando novamente $\sigma_1 = \sigma_2 \in \{1, 10, 100\}$ e agora $\mu_\ell = 1 + 100 \cdot i; i = 0, 1, 2, \dots, 100, \ell = 1, 2$. Deste modo obtivemos as figuras 2.4, 2.5 e 2.6.

Figura 2.4: Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 1$; $\mu_\ell = 1 + 100.i$; $i = 0, 1, 2, \dots, 100$; $\ell = 1, 2$

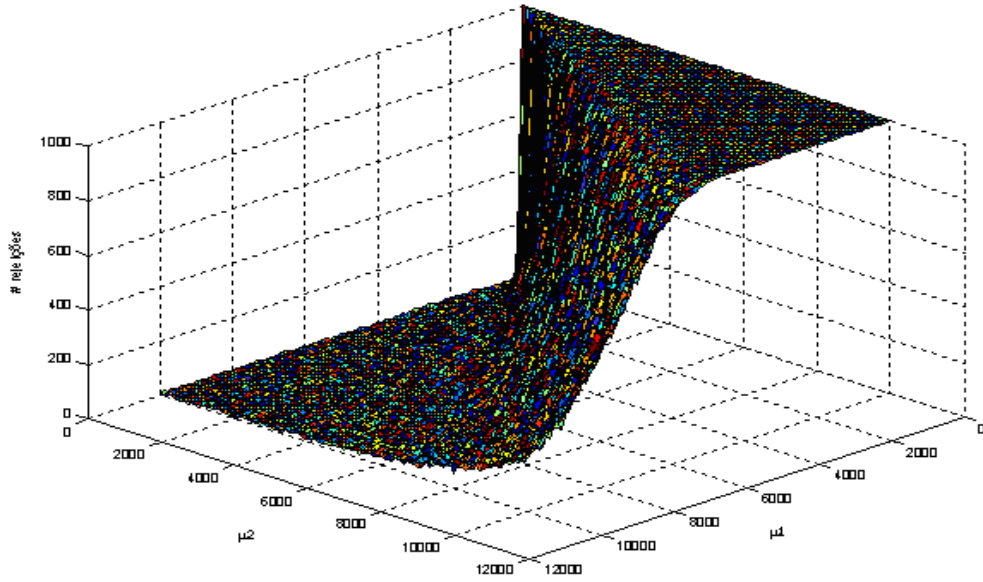


Figura 2.5: Número de rejeições para ($\alpha = 1\%$); $\sigma_1 = \sigma_2 = 10$; $\mu_\ell = 1 + 100.i$; $i = 0, 1, 2, \dots, 100$; $\ell = 1, 2$

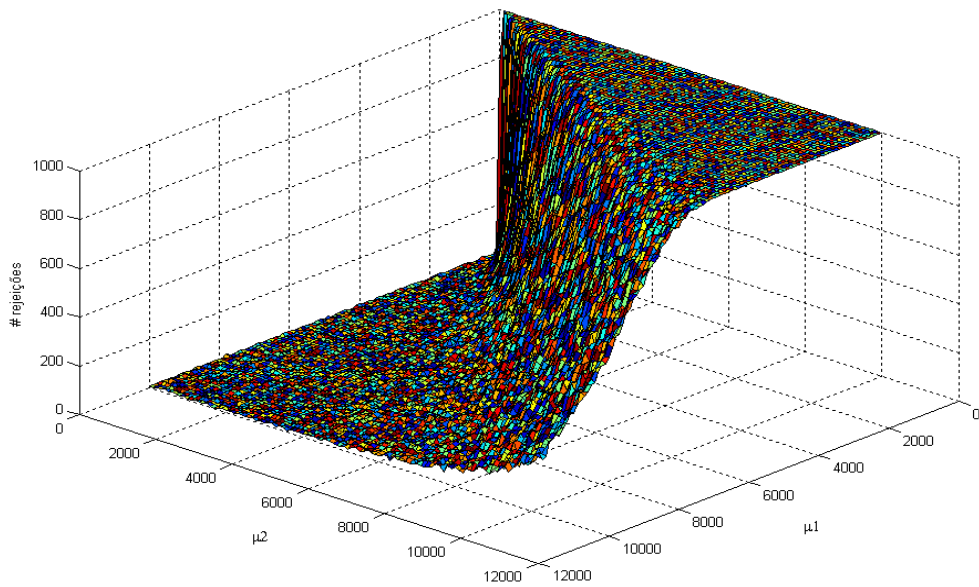
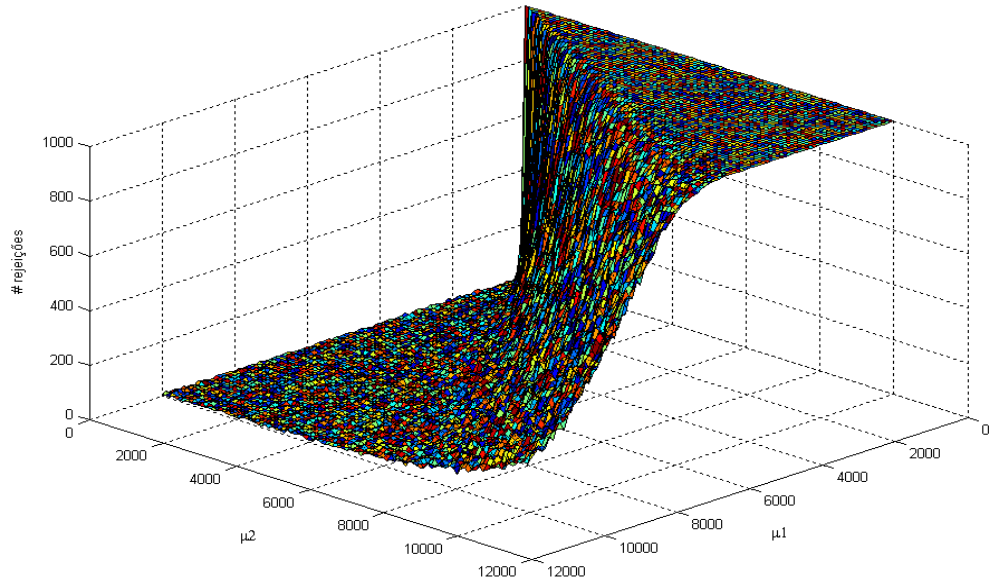


Figura 2.6: Número de rejeições para $(\alpha = 1\%); \sigma_1 = \sigma_2 = 100; \mu_\ell = 1 + 100.i; i = 0, 1, 2, \dots, 100; \ell = 1, 2$



As figuras sugerem a existência de dois valores θ' e θ'' tais que quando $\frac{\mu_1}{\mu_2} < \theta'$ [$> \theta''$] o rácio das rejeições da normalidade é muito alto [baixo]. A queda no rácio das rejeições ocorre para valores intermédios de $\theta = \frac{\mu_1}{\mu_2}$. Os valores de θ'' tendem a aproximar-se de 1.

Capítulo 3

Análise Conjunta de Regressões

3.1 Enquadramento histórico da ACR

Aastveit e Mejza (1992) ao considerarem os resultados estatísticos para estudo da interacção genótipo x ambiente colocaram a Análise Conjunta de Regressões, ACR no grupo dos métodos baseados na Análise de Regressões.

Históricamente a análise de regressões tem sido utilizada para interpretar redes de ensaios destinados à comparação de cultivares. O nome da técnica provinha de se ajustar uma regressão linear por cultivar. A técnica foi originalmente descrita por Mooers (1921) tendo sido introduzida por Yates & Cochran (1938), que calcularam a média total de todos os cultivares em determinado ambiente e designaram essa medida por índice ambiental. Os autores estudavam o comportamento dum cultivar através da regressão linear das produções do mesmo nos índices ambientais.

Na década de sessenta sucessivas investigações em ACR conduzem a grandes avanços da técnica, permitindo a sua utilização na comparação do desempenho de cultivares quando sujeitos a um vasto leque de capacidades produtivas dos pares (local, ano). Salientam-se os contributos dos trabalhos de Finlay & Wilkinson (1963), de Eberhart & Russel (1966) e de Perkins & Jinks (1968).

Apesar das inúmeras vantagens apresentadas pelo uso das técnicas de ACR, estas têm algumas limitações, por exemplo as apontadas no trabalho de Romagosa & Fox (1993), que salientam a fragilidade da ACR quando os ambientes testados são restritos em número e em diversidade, quando os cultivares incluídos foram já alvo de forte selecção ou quando não são suficientemente representativos para adequação ao modelo.

Wescott (1986) e Lin *et al.* (1986) criticam também o facto de a ACR não consid-

er ar variáveis ambientais específicas, e alertam para o facto de que com a escolha da média de todos os cultivares em cada condição como índice ambiental, poder não de verificar a independência entre as variáveis especialmente se o número de cultivares for inferior a quinze. No entanto outros autores contra-argumentaram e consideram essa crítica pouco relevante, por exemplo Becker & Leon (1988).

De facto a técnica ACR tem-se revelado extremamente útil, quer no que respeita à análise da interacção cultivar versus ambiente relativamente à produtividade ou rendimento, quer na abordagem aos problemas de estabilidade. A questão da estabilidade foi já bastante estudada e de entre os vários métodos desenvolvidos nesta área destaca-se o método de Toler (1990). Com base num modelo de regressão não linear nos parâmetros, este método surge como uma das alternativas para contornar os problemas de estimação do índice ambiental, permitindo também a realização de inferência no que respeita aos parâmetros, sendo posteriormente desenvolvido em Toler & Burrows (1998). Fazemos ainda referência a outros importantes trabalhos: Nguyen *et al.* (1980), Gray, E. (1982), Zobel *et al.* (1988), Jalalaluddin, M.D. e Harrison, S.A. (1993), Brancourt-Humel, M., Lecomte, C. (1994). Embora o problema da estabilidade seja extremamente importante, actualmente reveste-se de ênfase especial a procura de cultivares com boas capacidades de produção sendo neste aspecto que concentraremos o nosso estudo.

Em Portugal a investigação na área da ACR começa a evidenciar-se nos anos oitenta, com os trabalhos de Gusmão(1985), (1986a), (1986b). Entre outros resultados, o autor mostra que a consideração de índices ambientais distintos para os diferentes blocos, em vez de um único índice ambiental por ensaio, conduz a uma forte melhoria da precisão dos ajustamentos. São exemplos de outros trabalhos relevantes no âmbito da investigação em Portugal: Gusmão *et al.* (1991a), (1991b), (1992a), (1992b), (1993a), (1993b), (1995), Ferreira *et al.* (1990), Maçãs *et al.* (1990), Mexia *et al.* (1991) e Dias (2000).

Os trabalhos inicialmente apresentados por Gusmão (1985) e (1986a) reflectiam aplicações da ACR a redes de ensaios em blocos casualizados, em que os índices ambientais correspondiam aos rendimentos médios dos diversos blocos. Assim, para cada bloco era calculada a média das produções ajustando-se, cultivar a cultivar, regressões lineares das produções nas médias. Uma limitação desta técnica surge pelo facto de a sua aplicação apenas ser possível em situações de blocos completos casualizados e para além disso surgem algumas objecções quanto à consideração das produções médias como valores de uma variável controlada.

Estas limitações foram ultrapassadas no trabalho de Mexia *et al.* (1999), com a introdução dos índices ambientais L_2 e extensão das técnicas de ACR aos planos em blocos incompletos. Na dissertação de Pereira (2004) desenvolve-se de forma sistemática a ACR utilizando índices ambientais L_2 . A aplicação destes índices aos blocos incompletos é concretizada por um algoritmo e a sua utilização para além de permitir obter estimadores de máxima verosimilhança dos índices ambientais, permite a obtenção destes estimadores para outros parâmetros e a realização de testes de razão de verosimilhanças. Desta forma consegue-se a integração da ACR na inferência estatística. Em Pinto (2006) é feita uma abordagem da ACR em planos de melhoramento de trigo mole em Portugal.

No âmbito da investigação internacional note-se que em Digby (1979) é apresentado um método numérico iterativo para ajustamento de um modelo de regressão conjunta, para o caso de uma tabela incompleta de dupla entrada sem repetições, sem no entanto serem investigadas as propriedades estatísticas dos estimadores obtidos. No trabalho de Ng & Grunwald (1997) são também desenvolvidos métodos iterativos para o ajustamento conjunto de regressões e em Ng & Williams (2001) é feita uma comparação dos dois métodos - o método de Ng & Grunwald (1997) revela-se francamente vantajoso no que concerne à velocidade de convergência, para além da vantagem de permitir a realização de inferência.

Recentemente, Pinto(2006) mostrou como utilizar a ACR na condução de planos de melhoramento. Aplicou os seus resultados ao plano português de melhoramento de trigo mole (1986-2000).

Seguiremos este trabalho na medida que aplicaremos os nossos resultados ao plano português de melhoramento da aveia. Como veremos adiante completamos a teoria da ACR de forma a podermos incorporar informação adicional para além dos rendimentos. No entanto neste capítulo limitar-nos-emos a apresentar a ACR "clássica".

3.2 Considerações prévias

Abordaremos de seguida o caso geral e o caso completo. No caso geral não é exigido que todos os cultivares estejam presentes em todos os blocos. Nesse caso, as observações serão afectadas por pesos $p_{i,j}$ atribuídos ao j -ésimo cultivar, $j=1,\dots,J$, no i -ésimo bloco, $i=1,\dots,b$. Duma forma simplificada os pesos poderão assumir os valores 0[1] respectivamente quando o j -ésimo cultivar esteja ausente[presente] no i -ésimo bloco, sendo este o critério que adoptaremos nas abordagens efectuadas.

No caso completo teremos todos os cultivares presentes em todos os blocos, verificando-se $p_{i,1} = p_{i,2} = \dots = p_{i,J} = p_i$, $i=1, \dots, b$.

Estes pesos são usados para se ter em consideração a importância agronômica do local, tendo sido introduzidos por Dias (2000).

Na ausência dessa informação toma-se $p_i = 1$, $i=1, \dots, b$. Noutros casos, a colaboração com os técnicos agrônomos deve permitir a atribuição de pesos mais específicos que traduzam a importância dos blocos. Admitem-se desde já os pressupostos de normalidade, independência e homocedasticidade para os vectores das produções que consideramos.

Em qualquer dos casos, geral ou completo, o modelo será

$$E(y_{i,j}) = \alpha_j + \beta_j x_i, i = 1, \dots, b, j = 1, \dots, J \quad (3.2.1)$$

onde se pretende estimar os parâmetros

$$\boldsymbol{\alpha}_J = (\alpha_1, \dots, \alpha_J), \boldsymbol{\beta}_J = (\beta_1, \dots, \beta_J), \mathbf{x}_b = (x_1, \dots, x_b). \quad (3.2.2)$$

3.3 Variáveis não observáveis

O problema das variáveis não observáveis e das variáveis erroneamente observadas reveste-se de grande importância como nova área de investigação estatística. Trata-se de variáveis para as quais não se dispõe à partida de qualquer medida. Esta é uma situação distinta daquela em que se têm medidas sujeitas a erro. No nosso estudo surge-nos uma variável não observável: o índice ambiental. Esta variável controlada, mede, para cada par (local, ano), a respectiva capacidade produtiva. Sendo estes valores desconhecidos, terão de ser estimados, o que faremos através do algoritmo zig-zag, de que falaremos mais tarde.

3.4 Caso geral

Vamos ver como realizar o ajustamento começando por estudar a função objectivo que pretendemos minimizar. Os argumentos desta função serão os índices ambientais e os coeficientes das regressões. Em seguida, apresentaremos o algoritmo zig-zag. Este algoritmo é iterativo. Observe-se que o estudo que faremos da função objectivo permite, no fim de cada iteração, transformar os índices ambientais ajustados de maneira a manter invariante o seu intervalo de variação.

O algoritmo zig-zag foi construído para realizar o ajustamento em redes de ensaios conectadas. Mais adiante apresentaremos esse conceito.

3.4.1 Função Objectivo

Tal como referimos atrás tomemos $p_{i,j} = 0$ [1] se no i -ésimo bloco não estiver [estiver] presente o j -ésimo cultivar procurando-se minimizar

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2 \quad (3.4.3)$$

sendo $y_{i,j}$ a produção do j -ésimo cultivar no i -ésimo bloco, caso o cultivar esteja presente.

Quando o cultivar j não está presente no bloco i , tem-se como se referiu $p_{i,j} = 0$, significando que $y_{i,j}$ poderá assumir qualquer valor.

A escolha dos valores iniciais para as componentes do vector \mathbf{x}_b , no caso em que o delineamento é do tipo α -design, descrito no Apêndice A, pode ser feita atribuindo-se a cada bloco o valor do rendimento médio para o super-bloco em que o bloco está integrado. Sendo $x_i^0, i=1, \dots, b$, os valores iniciais para os índices ambientais, iniciamos com a minimização de

$$S(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i^0)^2 \quad (3.4.4)$$

em ordem aos vectores $\boldsymbol{\alpha}_J = (\alpha_1, \dots, \alpha_J)$ e $\boldsymbol{\beta}_J = (\beta_1, \dots, \beta_J)$.

Consideremos agora os vectores $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ pertencentes a $\mathbb{R}_b \times \mathbb{R}_J \times \mathbb{R}_J$ e estabeleçamos

Lema 3.1 *Sendo $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J)$, quando existirem ε e $\delta \neq \emptyset$ tal que*

$$\begin{cases} \mathbf{x}'_b &= \delta \mathbf{x}_b + \varepsilon \mathbf{1}_b \\ \boldsymbol{\alpha}'_J &= \boldsymbol{\alpha}_J - \frac{\varepsilon}{\delta} \boldsymbol{\beta}_J \\ \boldsymbol{\beta}'_J &= \frac{1}{\delta} \boldsymbol{\beta}_J \end{cases} \quad (3.4.5)$$

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

Dem: Considerando o caso em que $\delta = 1$ e $\varepsilon = 0$ teremos para todo o trio $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$, e portanto τ goza da propriedade reflexiva. Admitindo agora que $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J)$, e existindo $\delta \neq \emptyset$ e ε tais que ?? se verifica, então sendo $\delta' = \delta^{-1}$ e $\varepsilon' = -\frac{\varepsilon}{\delta}$ tem-se

$$\begin{cases} \mathbf{x}_b &= \delta' \mathbf{x}'_b + \varepsilon' \mathbf{1}_b \\ \boldsymbol{\alpha}_J &= \boldsymbol{\alpha}'_J - \frac{\varepsilon'}{\delta'} \boldsymbol{\beta}'_J \\ \boldsymbol{\beta}_J &= \frac{1}{\delta'} \boldsymbol{\beta}'_J \end{cases} \quad (3.4.6)$$

concluindo-se que τ goza da propriedade simétrica. Por último, admitamos que $(\mathbf{x}_b^i, \boldsymbol{\alpha}_J^i, \boldsymbol{\beta}_J^i) \tau (\mathbf{x}_b^{i+1}, \boldsymbol{\alpha}_J^{i+1}, \boldsymbol{\beta}_J^{i+1})$, existindo $\delta_i \neq 0$ e ε_i , tais que (??) se verifica quando $i = 1, 2$. Sendo agora, $\delta = \delta_1 \delta_2 \neq 0$ e $\varepsilon = \varepsilon_1 \delta_2 + \varepsilon_2$, as expressões (??) verificam-se para o par $[(\mathbf{x}_b^1, \boldsymbol{\alpha}_J^1, \boldsymbol{\beta}_J^1), (\mathbf{x}_b^3, \boldsymbol{\alpha}_J^3, \boldsymbol{\beta}_J^3)]$ o que mostra que τ goza da propriedade transitiva, ficando assim completa a demonstração. ■

O estabelecimento desta relação de equivalência é importante na medida em que a função objectivo toma valor constante para todos os trios de vectores numa classe de equivalência τ .

Consideremos agora

- $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\tau$ a classe de equivalência τ a que $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ pertence;
- $x_0 = \frac{1}{b} \sum_{i=1}^b x_i$;
- $D(\mathbf{x}_b) = \sqrt{\sum_{i=1}^b (x_i - x_0)^2}$

Teremos $D(\mathbf{x}_b) = 0$ se e só se todas as componentes do vector \mathbf{x}_b forem iguais, ou seja, quando \mathbf{x}_b pertencer ao espaço imagem $R(\mathbf{1}_b)$ da matriz coluna $\mathbf{1}_b$.

Estabeleçamos agora o

Lema 3.2 *Se $\mathbf{x}_b \in R(\mathbf{1}_b)$, $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\tau$ não tem nenhum trio de vectores em $\mathbf{V} = (\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) : (x_0 = 0) \wedge (\|\mathbf{x}_b\| = 1)$ mas se $\mathbf{x}_b \notin R(\mathbf{1}_b)$, $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\tau$ terá dois trios de vectores em \mathbf{V} .*

Dem: Se $\mathbf{x}_b \in R(\mathbf{1}_b)$ temos $x_1 = x_2 = \dots = x_b = x_0 = 0$, pelo que quando $x_0 = 0$ temos $\mathbf{x}_b = \mathbf{0}_b$ e desse modo $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]$ não pode ter nenhum trio em \mathbf{V} .

Por outro lado se $\mathbf{x}_b \notin R(\mathbf{1}_b)$ e $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J)$ tem-se $\mathbf{x}'_b = \varepsilon \mathbf{1}_b + \delta \mathbf{x}_b$ de forma que $x'_0 = \frac{1}{b} \sum_{i=1}^b x'_i = \delta x_0 + \varepsilon$ e $D(\mathbf{x}'_b) = |\delta| D(\mathbf{x}_b)$.

Para $(\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J) \in \mathbf{V}$, é necessário e suficiente que $(x'_0 = 0 \text{ e } D(\mathbf{x}'_b) = 1)$, o que acontece se e só se $(\delta = \pm D(x_b)^{-1} \text{ e } \varepsilon = -\delta x_0)$ e portanto $[(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)]_\tau$ conterà dois trios em \mathbf{V} . ■

Segue-se a

Proposição 3.1 *A função objectivo*

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2 \quad (3.4.7)$$

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

Dem: Quando $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J)$, verifica-se que $y_{i,j} - \alpha_j - \beta_j x_i = y_{i,j} - \alpha'_j - \beta'_j x'_i$, $i = 1, \dots, b$, $j = 1, \dots, J$, pelo que $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = S(\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J)$ e portanto a função objectivo toma valor constante nas classes de equivalência τ e de acordo com o lema 3.2, toma todos os seus valores, para $\mathbf{x}_b \notin R(\mathbf{1}_b)$, em \mathbf{V} . Se $\mathbf{x}_b \in R(\mathbf{1}_b)$, com $\delta = 1$ e $\varepsilon = -x_0$, teremos $\mathbf{x}'_b = \mathbf{1}\mathbf{x}_b - x_0\mathbf{1}_b = \mathbf{0}_b$, $\boldsymbol{\alpha}'_J = \boldsymbol{\alpha}_J - x_0\boldsymbol{\beta}_J$ e $\boldsymbol{\beta}'_J = \boldsymbol{\beta}_J$. Agora com $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \in \mathbf{V}$, temos $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \geq S(\mathbf{x}_b, \tilde{\boldsymbol{\alpha}}_J(\mathbf{x}_b), \tilde{\boldsymbol{\beta}}_J(\mathbf{x}_b))$ com $\tilde{\boldsymbol{\alpha}}_J(\mathbf{x}''_b)$ e $\tilde{\boldsymbol{\beta}}_J(\mathbf{x}''_b)$ estimadores de mínimos quadrados dos vectores de coeficientes. O mínimo da função objectivo a existir corresponde ao mínimo absoluto de $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ para $\mathbf{x}_b \notin R(\mathbf{1}_b)$.

Provemos agora que o mínimo absoluto existe, fazendo $\tilde{S}(\mathbf{x}_b) = S(\mathbf{x}_b, \tilde{\boldsymbol{\alpha}}_J(\mathbf{x}_b), \tilde{\boldsymbol{\beta}}_J(\mathbf{x}_b))$ tem-se

$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \geq \tilde{S}(\mathbf{x}_b)$, pelo que bastará mostrar que existe um mínimo absoluto para $\tilde{S}(\mathbf{x}_b)$, com $\mathbf{x}_b \notin R(\mathbf{1}_b)$. Utilizando raciocínio análogo ao utilizado para obter 3.4.6 obtém-se com $\delta \neq 0$, $\tilde{\boldsymbol{\alpha}}_J(\delta\mathbf{x}_b + \varepsilon\mathbf{1}_b) = \tilde{\boldsymbol{\alpha}}_J(\mathbf{x}_b) + \frac{\varepsilon}{\delta}\tilde{\boldsymbol{\beta}}_J(\mathbf{x}_b)$ e $\tilde{\boldsymbol{\beta}}_J(\delta\mathbf{x}_b + \varepsilon\mathbf{1}_b) = \frac{1}{\delta}\tilde{\boldsymbol{\beta}}_J(\mathbf{x}_b)$, vindo $\tilde{S}(\delta\mathbf{x}_b + \varepsilon\mathbf{1}_b) \in A$ com $A = \{\mathbf{x}_b : (x_0 = 0) \wedge (\|\mathbf{x}_b\| = 1)\}$.

Deste modo, $\tilde{S}(\mathbf{x}_b)$ tomará todos os seus valores para $\mathbf{x}_b \notin R(\mathbf{1}_b)$, em A . Como A é compacto, pelo teorema de Weirstrass, vê-se 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 $\mathbf{x}_b \in R(\mathbf{1}_b)$, o qual será o mínimo absoluto de $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$. ■

Continuando com o estudo da função objectivo observe-se que o gradiente e a matriz hessiana de S serão

$$\text{grad}(S) = \begin{bmatrix} \frac{\partial S}{\partial x_1} \\ \vdots \\ \frac{\partial S}{\partial x_b} \\ \frac{\partial S}{\partial \alpha_1} \\ \vdots \\ \frac{\partial S}{\partial \alpha_J} \\ \frac{\partial S}{\partial \beta_1} \\ \vdots \\ \frac{\partial S}{\partial \beta_J} \end{bmatrix} = -2 \begin{bmatrix} \sum_{j=1}^J p_{1,j}(y_{1,j} - \alpha_j - \beta_j x_1) \beta_j \\ \vdots \\ \sum_{j=1}^J p_{b,j}(y_{b,j} - \alpha_j - \beta_j x_b) \beta_j \\ \sum_{i=1}^b p_{i,1}(y_{i,1} - \alpha_1 - \beta_1 x_i) \\ \vdots \\ \sum_{i=1}^b p_{i,J}(y_{i,J} - \alpha_J - \beta_J x_i) \\ \sum_{i=1}^b p_{i,1}(y_{i,1} - \alpha_1 - \beta_1 x_i) x_i \\ \vdots \\ \sum_{i=1}^b p_{i,J}(y_{i,J} - \alpha_J - \beta_J x_i) x_i \end{bmatrix} \quad (3.4.8)$$

e

$$\text{Hess}(S) = \begin{bmatrix} \frac{\partial^2 S}{\partial x_1^2} & \cdots & \frac{\partial^2 S}{\partial x_1 \partial x_b} & \frac{\partial^2 S}{\partial x_1 \partial \alpha_1} & \cdots & \frac{\partial^2 S}{\partial x_1 \partial \alpha_J} & \frac{\partial^2 S}{\partial x_1 \partial \beta_1} & \cdots & \frac{\partial^2 S}{\partial x_1 \partial \beta_J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 S}{\partial x_b \partial x_1} & \cdots & \frac{\partial^2 S}{\partial x_b^2} & \frac{\partial^2 S}{\partial x_b \partial \alpha_1} & \cdots & \frac{\partial^2 S}{\partial x_b \partial \alpha_J} & \frac{\partial^2 S}{\partial x_b \partial \beta_1} & \cdots & \frac{\partial^2 S}{\partial x_b \partial \beta_J} \\ \frac{\partial^2 S}{\partial \alpha_1 \partial x_1} & \cdots & \frac{\partial^2 S}{\partial \alpha_1 \partial x_b} & \frac{\partial^2 S}{\partial \alpha_1^2} & \cdots & \frac{\partial^2 S}{\partial \alpha_1 \partial \alpha_J} & \frac{\partial^2 S}{\partial \alpha_1 \partial \beta_1} & \cdots & \frac{\partial^2 S}{\partial \alpha_1 \partial \beta_J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 S}{\partial \alpha_J \partial x_1} & \cdots & \frac{\partial^2 S}{\partial \alpha_J \partial x_b} & \frac{\partial^2 S}{\partial \alpha_J \partial \alpha_1} & \cdots & \frac{\partial^2 S}{\partial \alpha_J^2} & \frac{\partial^2 S}{\partial \alpha_J \partial \beta_1} & \cdots & \frac{\partial^2 S}{\partial \alpha_J \partial \beta_J} \\ \frac{\partial^2 S}{\partial \beta_1 \partial x_1} & \cdots & \frac{\partial^2 S}{\partial \beta_1 \partial x_b} & \frac{\partial^2 S}{\partial \beta_1 \partial \alpha_1} & \cdots & \frac{\partial^2 S}{\partial \beta_1 \partial \alpha_J} & \frac{\partial^2 S}{\partial \beta_1^2} & \cdots & \frac{\partial^2 S}{\partial \beta_1 \partial \beta_J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 S}{\partial \beta_J \partial x_1} & \cdots & \frac{\partial^2 S}{\partial \beta_J \partial x_b} & \frac{\partial^2 S}{\partial \beta_J \partial \alpha_1} & \cdots & \frac{\partial^2 S}{\partial \beta_J \partial \alpha_J} & \frac{\partial^2 S}{\partial \beta_J \partial \beta_1} & \cdots & \frac{\partial^2 S}{\partial \beta_J^2} \end{bmatrix}$$

$$= \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

onde

$$S_{11} = 2 \begin{bmatrix} \sum_{j=1}^J p_{1,j} \beta_j^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{j=1}^J p_{b,j} \beta_j^2 \end{bmatrix}, \quad (3.4.9)$$

$$S_{22} = 2 \begin{bmatrix} \sum_{i=1}^b p_{i,1} & \cdots & 0 & \sum_{i=1}^b p_{i,1} x_i & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{i=1}^b p_{i,j} & 0 & \cdots & \sum_{i=1}^b p_{i,j} x_i \\ \sum_{i=1}^b p_{i,1} x_i & \cdots & 0 & \sum_{i=1}^b p_{i,1} x_i^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{i=1}^b p_{i,j} x_i & 0 & \cdots & \sum_{i=1}^b p_{i,j} x_i^2 \end{bmatrix}, \quad (3.4.10)$$

$$S_{12} = \begin{bmatrix} p_{11} \beta_1 & \cdots & p_{1J} \beta_J & p_{1,1} [x_1 \beta_1 - (y_{1,1} - \alpha_1 - \beta_1 x_1)] & \cdots & p_{1,J} [x_1 \beta_J - (y_{1,J} - \alpha_J - \beta_J x_1)] \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{b,1} \beta_1 & \cdots & p_{b,J} \beta_J & p_{b,1} [x_b \beta_1 - (y_{b,1} - \alpha_1 - \beta_1 x_b)] & \cdots & p_{b,J} [x_b \beta_J - (y_{b,J} - \alpha_J - \beta_J x_b)] \end{bmatrix}$$

e

$$S_{21} = (S_{12})' \quad (3.4.11)$$

Proposição 3.2 *As matrizes $Hess(S_{11})$ e $Hess(S_{22})$, são definidas positivas.*

Dem: Relativamente à matriz $Hess(S_{11})$ temos a forma quadrática

$$\frac{1}{2}[g_1 \cdots g_b]Hess(S_{11})[g_1 \cdots g_b]' = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} g_i^2 \beta_j^2 > 0 \quad (3.4.12)$$

para qualquer vector não nulo $[g_1 \cdots g_b]'$ e logo $Hess(S_{11})$ é definida positiva.

Para a matriz $Hess(S_{22})$ temos a correspondente forma quadrática

$$\begin{aligned} \frac{1}{2}[g_1 \cdots g_J t_1 \cdots t_J]2Hess(S_{22})[g_1 \cdots g_J t_1 \cdots t_J] &= \sum_{j=1}^J (g_j^2 \sum_{i=1}^b p_{i,j} + 2g_j t_j \sum_{i=1}^b p_{i,j} x_i + t_j^2 \sum_{i=1}^b p_{i,j} x_i^2) \\ &= \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (g_j^2 + 2x_i g_j t_j + x_i^2 t_j^2) \\ &= \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (g_j + x_i t_j)^2 > 0 \end{aligned} \quad (3.4.13)$$

para qualquer vector $[g_1 \cdots g_J t_1 \cdots t_J]'$ não nulo, o que completa a demonstração. ■

Pereira (2004) e Pinto (2006), apresentam uma simplificação para se poder aproximar a matriz hessiana de S pela matriz

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

Esta simplificação tem por base a suposição de que após algumas iterações do algoritmo zig-zag se obtém um bom ajustamento, sendo obtidos os estimadores dos índices ambientais \tilde{x}_i , $i = 1, \dots, b$ e dos coeficientes de regressão ajustados $(\tilde{\alpha}_j, \tilde{\beta}_j)$, $j = 1, \dots, J$.

A boa qualidade do ajustamento permitirá concluir que os valores dos resíduos

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

serão próximos de zero.

Podemos então reescrever as derivadas parciais $\frac{\partial^2 S}{\partial x_i \partial \beta_j}$ de forma bem mais simples, vindo

$$\frac{\partial^2 S}{\partial x_i \partial \beta_j} = -2p_{i,j}(y_{i,j} - \tilde{\alpha}_j - \tilde{\beta}_j \tilde{x}_i) + 2p_{i,j} \tilde{\beta}_j x_i \approx 2p_{i,j} \tilde{\beta}_j \tilde{x}_i$$

Demonstra-se de seguida que a matriz D é definida positiva, sendo de prever um bom desempenho do método. Sendo $\mathbf{u}'_{b+2J} = [\mathbf{g}'_b, \mathbf{t}'_J, \mathbf{s}'_J]$, virá a forma quadrática

$$\begin{aligned} \mathbf{u}'_{b+2J} \mathbf{D} \mathbf{u}_{b+2J} &= 2 \left[\sum_{i=1}^b \left(\sum_{j=1}^J p_{i,j} \tilde{\beta}_j^2 \right) g_i^2 + \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \right) t_j^2 + \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \tilde{x}_i^2 \right) s_j^2 \right] \\ &+ 2 \times 2 \left[\sum_{i=1}^b \sum_{j=1}^J p_{i,j} \tilde{\beta}_j \tilde{x}_i g_i s_j + \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \tilde{\beta}_j \tilde{x}_i g_i t_j + \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \tilde{x}_i \right) t_j s_j \right] \\ &= 2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} [\tilde{\beta}_j^2 g_i^2 + 2\tilde{\beta}_j g_i (t_j + s_j \tilde{x}_i) + t_j^2 + 2t_j s_j \tilde{x}_i + s_j^2 \tilde{x}_i^2] \\ &= 2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} [\tilde{\beta}_j g_i + (t_j + s_j \tilde{x}_i)]^2 > 0 \end{aligned}$$

concluindo-se que numa vizinhança do óptimo podemos considerar a matriz *Hess* (S) definida positiva e nesse caso será de esperar um bom comportamento do algoritmo zig-zag.

3.4.2 Redes Conectadas

Uma rede de ensaios pode ser considerada como um único ensaio, o qual pode ser conectado ou não. No caso do ensaio não ser conectado pode ser decomposto em sub-ensaios conectados e podemos dizer que a rede de ensaios foi decomposta em sub-redes conectadas.

Na minimização descrita no algoritmo zig-zag, obtém-se uma sequência de mínimos, existindo um mínimo absoluto se a rede for conectada. Vamos agora seguir Pereira (2004) e Pinto (2006) para estabelecer a

Proposição 3.3 *Quando a rede de ensaios é não conectada, os conjuntos de cultivares presentes nas várias sub-redes conectadas, são disjuntos, existindo mínimos absolutos para a função objectivo que não são equivalentes τ .*

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 Kageyama (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. Sejam A_ℓ , $\ell = 1, \dots, L$, os conjuntos dos índices dos blocos para as sub-redes conectadas, com $L \geq 2$, nas quais a rede de ensaios se divide.

Teremos então $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{\ell=1}^L S_\ell(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ com

$$S_\ell(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{i \in A_\ell} \sum_{j=1}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2, \ell = 1, \dots, L$$

Uma vez verificar-se os blocos e os cultivares serem distintos para as diferentes sub-redes o mínimo de S será a soma dos mínimos para S_ℓ , $\ell = 1, \dots, L$.

Através da proposição 3.1 podemos observar que os mínimos para as S_ℓ , $\ell = 1, \dots, L$, podem ser escolhidos satisfazendo as condições $\sum_{i \in A_\ell} x_i = a_\ell$ e $\sum_{i \in A_\ell} x_i^2 =$

g_ℓ , $\ell = 1, \dots, L$, com $\sum_{\ell=1}^L a_\ell = 0$ e $\sum_{\ell=1}^L g_\ell = 1$.

Juntando estes mínimos parciais, obtemos mínimos para S que pertencem a \mathbf{V} , mas não são equivalentes τ o que estabelece a tese. ■

Conclui-se que, para se efectuar uma análise global duma rede de ensaios é exigível que a mesma seja conectada. Caso o não seja, deve proceder-se à sua decomposição em sub-redes conectadas maximais e proceder-se a uma análise individual.

3.4.3 Algoritmo zig-zag

O algoritmo zig-zag baseia-se num processo iterativo onde as minimizações alternam entre os coeficientes de regressão e os regressores não observáveis: os índices ambientais. Frequentemente é usado no ajustamento de regressões lineares dos rendimentos nos índices ambientais, apresentando em geral um bom desempenho. Neste ponto é feita a descrição do algoritmo e apresentadas as justificações que nos permitem garantir a boa qualidade dos ajustamentos.

Tal como vimos na secção 3.4.1 procuramos minimizar a função

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

onde $y_{i,j}$ é o rendimento do j -ésimo cultivar no i -ésimo bloco se $p_{i,j}$ e b o número de pares (local,ano). Para tal devemos começar por escolher valores iniciais para as componentes do vector de estrutura \mathbf{x}_b .

No caso de se ter ensaios com blocos completos, consideram-se os rendimentos médios dos blocos, como valores iniciais para os índices ambientais.

Se tivermos um delineamento do tipo α -design, ver Patterson e Williams(1976), os blocos estão agrupados em superblocos, onde cada cultivar estará representado α vezes. Nesta situação atribui-se a cada bloco, o rendimento médio para o super-bloco a que pertence, como valor inicial para o respectivo índice ambiental. Tomaremos a opção de considerar os rendimentos médios dos blocos como valores iniciais para os respectivos índices ambientais, e sendo x_i^0 , $i = 1, \dots, b$, os respectivos valores iniciais, começaremos por minimizar

$$S_1(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i^0)^2$$

em ordem aos vectores $\boldsymbol{\alpha}_J = (\alpha_1, \dots, \alpha_J), \boldsymbol{\beta}_J = (\beta_1, \dots, \beta_J)$.

Os estimadores usuais dos mínimos quadrados permitem obter o mínimo $(\tilde{\alpha}_j^1, \tilde{\beta}_j^1)$, $j = 1, \dots, J$. Os índices superiores significam que se trata dos resultados relativos à primeira iteração.

Seguidamente procede-se à minimização do vector \mathbf{x}_b , sendo neste caso a função objectivo

$$S_1(\mathbf{x}_b) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \tilde{\alpha}_j^1 - \tilde{\beta}_j^1 x_i^1)^2 = \sum_{i=1}^b h_i^1(x_i^1)$$

com

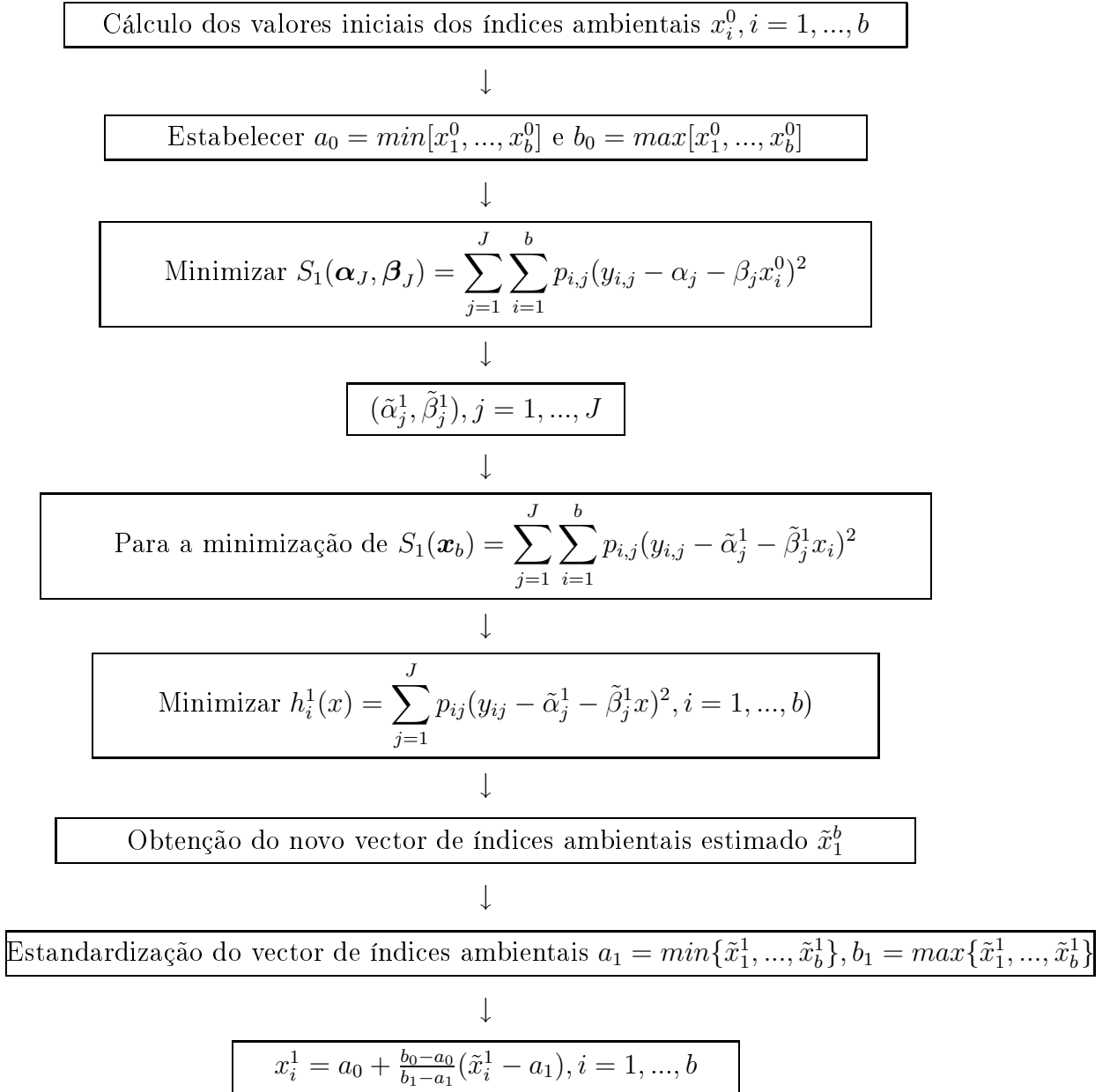
$$h_i^1(x) = \sum_{j=1}^J p_{i,j} (y_{i,j} - \tilde{\alpha}_j^1 - \tilde{\beta}_j^1 x)^2, \quad i = 1, \dots, b$$

sendo as componentes dos mínimos dadas por

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

repetindo-se as iterações até a soma de quadrados dos resíduos estabilizar. Segue-se uma descrição esquemática do algoritmo

Esquema 1 - Algoritmo Zig-Zag



Repete-se o procedimento até que as sucessivas somas das somas dos quadrados dos resíduos difiram a menos de uma constante suficientemente pequena previamente fixada.

Com a aplicação sucessiva do processo descrito, obtém-se uma sucessão decrescente. Verifica-se no entanto, tal com vimos atrás, que só existirá um mínimo absoluto de S se a rede de ensaios for conectada.

Em Pereira (2004) é realçada a importância da conexão da rede de ensaios por esta razão, e porque só assim se pode efectuar uma análise integrada da mesma. Acontece porém que mesmo para as redes de ensaios sejam conectadas, podemos ter mínimos absolutos S que não são equivalentes τ .

Para que os mínimos de S sejam equivalentes τ é necessário e suficiente, ver Pereira(2004), que

- i) A rede de ensaios seja conectada;
- ii) $J \leq \text{Int}(\frac{n-b}{2} + 1)$, onde n é o número total de parcelas existentes nos b blocos.

A partir daqui estamos nas condições para construir o algoritmo para obtermos o mínimo $(\tilde{x}_b, \tilde{\alpha}_J, \tilde{\beta}_J)$ da função objectivo $S(x_b, \alpha_J, \beta_J)$

3.4.4 Análise dos Resíduos

A Análise Conjunta de Regressões (ACR) é, como já referido atrás, uma técnica bem conhecida em problemas de comparação de cultivares. Os índices ambientais em ACR correspondem a um regressor não observável que mede a produtividade dos blocos nos campos de ensaio. Em geral o algoritmo zig-zag descrito no ponto anterior é usado no ajustamento. Neste algoritmo as minimizações para os coeficientes de regressão alternam com os adequados aos índices ambientais. O algoritmo revela um excelente desempenho, ver Mexia e Pinto(2004), Pereira(2004), Pereira e Mexia(2003), Pereira e Mexia(2003a) e Pinto(2006), embora se encontre ainda em aberto o problema da demonstração da sua convergência para o mínimo absoluto da soma dos quadrados dos resíduos. Apresentaremos uma validação dos ajustamentos usando o algoritmo zig-zag, através dum modelo linear para os resíduos, e respectiva discussão.

Modelo para os desvios

Pode pensar-se que se o ajustamento for defeituoso tal que, com (α_j, β_j) e $x_i[(\tilde{\alpha}_j, \tilde{\beta}_j)$ e $\tilde{x}_i]$ os exactos[ajustados] coeficientes de regressão e índices ambientais, teríamos

$$\begin{cases} \alpha_j = \tilde{\alpha}_j + \gamma_j \\ \beta_j = \tilde{\beta}_j + \eta_j \quad ; \quad j = 1, \dots, J \quad , \quad i = 1, \dots, b \\ x_i = \tilde{x}_i + u_i \end{cases} \quad (3.4.14)$$

onde γ_j, η_j e u_i são os desvios correspondentes respectivamente a α_j, β_j e x_i . Somos então conduzidos a testar a significância destes desvios. Para estabelecer estes testes consideremos o modelo para os resíduos de regressão

$$\hat{y}_{i,j} = y_{i,j} - (\tilde{\alpha}_j + \tilde{\beta}_j \tilde{x}_i) \quad (3.4.15)$$

Substituindo os $(\tilde{\alpha}_j, \tilde{\beta}_j)$ e \tilde{x}_i pelas suas expressões obtemos

$$\hat{y}_{i,j} = \gamma_j + \tilde{x}_i \eta_j + \tilde{\beta}_j u_i + \eta_j u_i \quad (3.4.16)$$

Assumindo que os termos $\eta_j u_i$, podem ser desprezados podemos reescrever o modelo na forma matricial

$$\underline{\hat{y}} = \overset{\circ}{\overset{\circ}{X}} \underline{\theta}' + e \quad (3.4.17)$$

onde

$$\begin{cases} \underline{\hat{y}} = [\hat{y}'_1 \dots \hat{y}'_J] \\ \overset{\circ}{\overset{\circ}{X}} = [I_J \otimes \overset{\circ}{\overset{\circ}{X}} \quad \underline{\tilde{\beta}} \otimes I_b] \\ \underline{\theta}' = (\gamma_1, \eta_1, \dots, \gamma_J, \eta_J, u_1, \dots, u_b) \end{cases} \quad (3.4.18)$$

e e representa o vector do erro associado a $y_{i,j}$, assumindo-se os usuais pressupostos.

Temos ainda

$$\overset{\circ}{\overset{\circ}{X}} = \begin{bmatrix} 1 & \tilde{x}_1 \\ \vdots & \vdots \\ 1 & \tilde{x}_b \end{bmatrix} \quad (3.4.19)$$

O estimador dos mínimos quadrados para $\underline{\theta}$ será

$$\tilde{\underline{\theta}} = (\overset{\circ}{\overset{\circ}{X}} \overset{\circ}{\overset{\circ}{X}})^{-1} * \overset{\circ}{\overset{\circ}{X}}' \underline{\hat{y}} \quad (3.4.20)$$

e a soma dos quadrados dos resíduos para o modelo ajustado dada por

$$SS_{modelo} = \underline{\hat{y}}' \underline{\hat{y}} - \underline{\hat{y}}' \overset{\circ}{\overset{\circ}{X}} \tilde{\underline{\theta}} \quad (3.4.21)$$

Para avaliarmos a qualidade do ajustamento obtida pelo algoritmo zig-zag podemos calcular

$$R^2 = 1 - (SS_{modelo}/SS_{zig-zag}) \quad (3.4.22)$$

onde $SS_{zig-zag}$ representa a soma dos quadrados dos resíduos originais obtidos pelo algoritmo zig-zag. Contrariamente ao que é usual estamos neste caso interessados em obter valores baixos de R^2 , pois isso significará um melhor ajustamento do modelo. Este resultado será indicador de que não se extraiu uma informação significativa pelo modelo ajustado para os resíduos. Esta falta de informação associada aos desvios valida o modelo inicial ajustado uma vez que ele teria retirado toda a informação relevante.

Aplicação

O caso estudado baseia-se em dados de produção de aveia, obtidos em ensaios realizados nos anos de 1993 a 1997, pela Estação Nacional de Melhoramento de Plantas, que gentilmente nos autorizou a sua análise. Nestas experiências, foram ensaiados, em cada local/ano o mesmo número de cultivares (11), distribuídos em cada local por um mesmo número de blocos (4). Convém salientar que, de ano para ano, os locais, o número de locais, os cultivares e o número de observações variam. Na tabela 3.1 indicam-se os locais e anos dos ensaios.

Tabela 3.1: Locais e anos dos ensaios

Local / Estação Experimental	Ano
Elvas(ENMP)	1993,94,95,97
Elvas(Herdade da Comenda)	1994,96,97
Benavila	1994
Évora	1993,94,95,96,97
Beja	1993,94,95,96,97
Pegões	1993,95,96

Aplicando o algoritmo zig-zag para cada ano e cada bloco obtemos usando (3.4.17), os coeficientes de regressão e os índices ambientais ajustados. Esses resultados estão nas Tabelas 3.2 e 3.3

Tabela 3.2: Coeficientes de regressão ajustados

<i>Cultivar</i>	1993		1994		1995		1996		1997	
	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$
AE8901	-108.46	1.270	-	-	-	-	-	-	-	-
AE9004	-137.07	1.205	207.10	0.928	-	-	-	-	-	-
AE8902	-37.35	1.026	-	-	-	-	-	-	-	-
AE9005	-7.76	1.01	-342.74	1.248	-	-	-	-	-	-
S. VICENTE	-8.26	0.988	-	-	-	-	-	-	-	-
AE9002	12.96	0.982	-231.79	1.11	67.51	0.972	-595.26	1.255	374.19	0.884
ST	49.60	0.95	196.52	0.886	32.11	1.004	679.48	0.906	220.15	1.128
ALEIXO										
AE9001	59.11	0.909	9.94	0.911	-	-	-	-	-	-
AE9003	67.56	0.897	13.94	1.005	-79.20	1.235	886.51	0.914	-186.81	1.457
AE8801	91.26	0.890	-	-	-	-	-	-	-	-
AVON	60.58	0.815	90.6	0.928	73.46	0.903	128.66	0.933	-69.67	0.919
AE9303	-	-	-122.87	1.082	0.12	1.01	67.49	0.970	-111.66	1.01
AE9101	-	-	-85.04	0.969	-249.91	1.210	63.97	0.947	-42.37	0.943
AE9302	-	-	170.78	0.948	-24.32	1.010	-197.36	1.117	-14.22	1.103
AE9301	-	-	168.33	0.936	130.79	0.910	835.84	0.728	234.15	0.682
AE9401	-	-	-	-	-72.59	1.052	-1079.9	1.263	-356.53	1.179
AE9402	-	-	-	-	68.19	0.872	-156.39	0.991	-32.65	0.891
AE9403	-	-	-	-	73.46	0.713	-268.83	0.824	-40.4	0.891

Tabela 3.3: Índices ambientais

Ano	1993	1994	1995	1996	1997
Número de blocos	16	20	16	16	16
	818.59	3073.93	1502.61	4353.88	1163.740
	681.89	3715.50	1357.96	4787.51	1512.78
	692.94	3793.97	1084.18	4835.70	1544.79
	684.85	3891.24	1347.66	4897.19	1306.06
	512.38	3376.61	1182.01	1976.45	1928.69
	583.62	3290.09	1452.43	3693.10	2061.29
	564.36	3934.71	1513.96	3638.71	2183.35
	606.74	3838.16	1465.87	3573.67	2384.54
	1362.70	4112.04	1394.12	2517.13	813.82
	1298.35	5032.46	1848.48	3087.93	995.58
	1325.97	5004.31	1881.31	3287.93	1090.89
	1539.50	4888.40	1680.99	3516.68	935.58
	134.34	642.85	614.40	1807.90	1450.49
	147.60	949.77	774.51	1988.30	2103.36
	161.45	847.68	389.83	1807.13	2535.30
	169.13	913.36	380.41	1634.20	2836.36
		3289.87			
		2976.09			
		2673.59			
		3492.01			

De seguida usamos os índices ambientais no ajustamento dos modelos. Na tabela seguinte apresentamos os coeficientes ajustados obtidos através do software MatLab.

Tabela 3.4: Coeficientes ajustados

1993			1994			1995			1996			1997		
γ_j	η_j	u_i	γ_j	η_j	u_i	γ_j	η_j	u_i	γ_j	η_j	u_i	γ_j	η_j	u_i
68.6	0.02	17.2	-10.8	0.02	-48.0	-3.0	-0.07	-39.7	-4.6	0.00	17.5	-17.6	-0.01	-7.1
-35.1	0.05	-36.6	-48.1	0.00	-84.8	-13.4	-0.05	-9.0	3.8	0.00	-2.8	4.9	0.00	17.6
7.9	0.07	21.8	4.4	0.00	-178.8	-0.6	-0.09	7.0	3.6	0.00	-12.3	-0.3	0.01	7.9
42.5	0.02	26.3	-40.8	0.00	-123.4	-76.4	-0.01	14.7	5.4	0.00	-13.8	-5.6	0.01	9.2
45.7	-0.02	-3.8	19.0	-0.02	-138.4	48.0	0.07	-10.1	6.7	0.00	1.8	-7.5	-0.01	24.2
34.3	0.01	-63.0	-3.7	0.01	-20.5	-1.7	-0.03	-44.2	10.3	0.00	-8.7	-25.3	0.01	-3.7
48.0	-0.02	17.8	-9.5	-0.01	-41.4	87.1	0.03	-12.9	-14.1	0.00	-6.1	-12.1	-0.03	-17.8
31.9	-0.01	11.1	-12.4	-0.01	142.5	28.6	0.03	20.0	13.9	0.01	4.0	7.9	0.00	-11.0
58.9	0.02	2.3	14.3	0.04	49.1	33.0	-0.06	-8.9	26.4	0.00	14.0	47.0	-0.03	0.00
-1.3	-0.05	5.5	-3.5	0.03	185.4	-56.3	0.04	-138.2	6.9	0.01	-22.5	-5.2	0.01	30.4
37.5	-0.08	-2.9	-27.9	-0.01	106.7	26.8	0.02	-54.8	3.6	0.00	-3.0	-0.6	0.00	5.7
		3.2			-175.1			-26.8			-17.0			-0.9
		-6.2			31.2			48.5			19.1			-6.0
		85.3			6.0			14.2			-1.3			-3.0
		28.4			30.7			-18.6			6.2			30.1
		-5.5			17.2			-37.7			-0.1			-26.0
					-0.6									
					18.8									
					7.4									
					-8.9									

Por último na tabela 3.5 apresentam-se a soma dos quadrados dos resíduos obtidos pelo zig-zag e pelo modelo e ainda os coeficientes R^2 . Os resultados obtidos evidenciam uma validação do algoritmo neste caso.

Tabela 3.5: Somas de quadrados dos resíduos e coeficientes R^2

	1993	1994	1995	1996	1997
<i>Soma de quadrados (zig - zag)</i>	$1.9283 * 10^6$	$5.4068 * 10^7$	$1.0998 * 10^7$	$3.1286 * 10^7$	$2.0115 * 10^7$
<i>Soma de quadrados (modelo)</i>	$1.9283 * 10^6$	$5.4067 * 10^7$	$1.0998 * 10^7$	$3.1285 * 10^7$	$2.0104 * 10^7$
R^2	$1.4 * 10^{-5}$	$1.8 * 10^{-5}$	$4.9 * 10^{-7}$	$2.6 * 10^{-5}$	$5.5 * 10^{-4}$

3.5 Caso completo

No início a ACR foi aplicada a redes de ensaios em blocos casualizados. Nesse contexto em Gusmão(1985), (1986a) e (1986b) são utilizados os índices ambientais correspondentes aos diferentes blocos medidos pelos respectivos rendimentos médios, mostrando-se que a precisão das análises dos delineamentos em blocos casualizados aumenta de modo significativo se forem considerados índices ambientais por bloco em vez de um só índice ambiental por ensaio, explicando-se que tal se deve ao facto de haver um aumento do número de pontos obtidos por regressão.

Neste caso a técnica clássica ajusta cultivar a cultivar, regressões lineares das produções nas médias obtidas para cada bloco. Este tipo de procedimento tem no entanto duas desvantagens:

- i) não ser de todo correcto considerar as produções médias como valores duma variável controlada;
- ii) ser aplicável apenas no caso completo.

Verifica-se no entanto, apesar de algumas dificuldades, que a técnica tem sido aplicada com bons resultados. Como alternativa pode aplicar-se o algoritmo zig-zag ou a técnica de dupla minimização que a seguir se descreve.

3.5.1 Dupla minimização

Abordaremos este problema considerando que todos os cultivares estão presentes em todos os blocos. Neste caso teremos a função objectivo

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

onde os pesos p_i correspondem a blocos do mesmo ensaio.

Consideremos agora as somas

$$\begin{aligned} T_x &= \sum_{i=1}^b p_i x_i & ; & & T_{jy} &= \sum_{i=1}^b p_i y_{i,j} & ; & & j &= 1, \dots, J \\ S_{xx} &= \sum_{i=1}^b p_i x_i^2 & ; & & S_{jxy} &= \sum_{i=1}^b p_i x_i y_{i,j} & ; & & S_{jyy} &= \sum_{i=1}^b p_i y_{i,j}^2 \end{aligned}$$

e as somas corrigidas

$$s_{xx} = S_{xx} - \frac{\left(\sum_{i=1}^b p_i x_i\right)^2}{\sum_{i=1}^b p_i} = \sum_{i=1}^b p_i \left(x_i - \frac{\sum_{i=1}^b p_i x_i}{\sum_{i=1}^b p_i}\right)^2$$

$$s_{jxy} = S_{jxy} - \frac{1}{\sum_{i=1}^b p_i} \sum_{i=1}^b p_i x_i \sum_{i=1}^b p_i y_{i,j} = \sum_{i=1}^b p_i \left(x_i - \frac{\sum_{i=1}^b p_i x_i}{\sum_{i=1}^b p_i}\right) \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i}\right) \quad ;$$

$$j = 1, \dots, J$$

$$s_{jyy} = S_{jyy} - \frac{1}{\sum_{i=1}^b p_i} \left(\sum_{i=1}^b p_i y_{i,j}\right)^2 = \sum_{i=1}^b p_i \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i}\right)^2 \quad ; \quad j = 1, \dots, J$$

Agora conhecendo-se o vector \mathbf{x}_b podemos obter os estimadores condicionais dos coeficientes

$$\tilde{\beta}_j(\mathbf{x}_b) = \frac{s_{jxy}}{s_{xx}} = \frac{\sum_{i=1}^b p_i x_i y_{i,j}}{\sum_{i=1}^b p_i x_i^2} \quad ; \quad j = 1, \dots, J$$

e

$$\tilde{\alpha}_j(\mathbf{x}_b) = \frac{1}{\sum_{i=1}^b p_i} (T_{jy} - \tilde{\beta}_j(\mathbf{x}_b) T_x) = \frac{1}{\sum_{i=1}^b p_i} \left(\sum_{i=1}^b p_i y_{i,j} - \tilde{\beta}_j(\mathbf{x}_b) \sum_{i=1}^b p_i x_i \right) \quad ;$$

$$j = 1, \dots, J$$

assim como o mínimo condicional para a soma dos quadrados dos resíduos

$$\bar{S}(\mathbf{x}_b) = S(\mathbf{x}_b, \tilde{\alpha}_J(\mathbf{x}_b), \tilde{\beta}_J(\mathbf{x}_b)) = \sum_{i=1}^b \sum_{j=1}^J p_i (y_{i,j} - \tilde{\alpha}_J(\mathbf{x}_b) - \tilde{\beta}_J(\mathbf{x}_b) x_i)^2 =$$

$$\sum_{j=1}^J s_{jyy} - \sum_{j=1}^J \frac{s_{jxy}^2}{s_{xx}}$$

Para minimizarmos $\bar{S}(\mathbf{x}_b)$, dado que $\sum_{j=1}^J s_{jyy}$ não depende de \mathbf{x}_b , temos que

$$\text{maximizar } \sum_{j=1}^J \frac{s_{jxy}^2}{s_{xx}}.$$

Quando se consideram as restrições $\sum_{i=1}^b p_i x_i = 0$ e $\sum_{i=1}^b p_i x_i^2 = 1$, virá $s_{xx} = 1$, pelo que o problema reduz-se à maximização

$$\begin{aligned} M(\mathbf{x}_b) &= \sum_{j=1}^J s_{jxy}^2 = \sum_{j=1}^J \left(\sum_{i=1}^b p_i \left(x_i - \frac{\sum_{i=1}^b p_i x_i}{\sum_{i=1}^b p_i} \right) \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i} \right) \right)^2 \\ &= \sum_{j=1}^J \left(\sum_{i=1}^b p_i x_i \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i} \right) \right)^2 \end{aligned}$$

verificando-se

$$\sum_{i=1}^b p_i \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i} \right) = 0 ; j = 1, \dots, J$$

e sendo \mathbf{Y}_j^{+b} o vector de componentes $p_i \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i} \right)$, $i = 1, \dots, b$, $j = 1, \dots, J$,

, temos

$$\begin{aligned} M(\mathbf{x}_b) &= \sum_{j=1}^J \left(\sum_{i=1}^b p_i x_i \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i} \right) \right)^2 = \sum_{j=1}^J (\mathbf{x}'_b \mathbf{Y}_j^{+b})^2 = \sum_{j=1}^J (\mathbf{x}'_b \mathbf{Y}_j^{+b}) (\mathbf{Y}_j^{+b} \mathbf{x}_b) \\ &= \sum_{j=1}^J \mathbf{x}'_b (\mathbf{Y}_j^{+b} \mathbf{Y}_j'^{+b}) \mathbf{x}_b = \mathbf{x}'_b \left(\sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j'^{+b} \right) \mathbf{x}_b \end{aligned}$$

Proposição 3.4 O vector \mathbf{x}_b que satisfaz as restrições $\sum_{i=1}^b p_i x_i = 0$ e $\sum_{i=1}^b p_i x_i^2 = 1$, maximiza $M(\mathbf{x}_b)$ é o primeiro vector próprio da matriz $\left(\sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j'^{+b} \right)$

Dem: Dado que $\sum_{i=1}^b p_i (y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{b}) = 0$; $j = 1, \dots, J$, então temos $(\sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j^{\prime +b}) (\frac{1}{b^{1/2}} \mathbf{1}_b) =$

$\sum_{j=1}^J \mathbf{Y}_j^{+b} (\mathbf{Y}_j^{\prime +b} \frac{1}{b^{1/2}} \mathbf{1}_b) = 0$, e logo $\frac{1}{b^{1/2}} \mathbf{1}_b$ será um vector próprio de $(\sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j^{\prime +b})$ estando-lhe associado um valor próprio nulo.

Agora tendo-se $u' \sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j^{\prime +b} u_b = \sum_{j=1}^J (u'_b \mathbf{Y}_j^{+b}) (\mathbf{Y}_j^{\prime +b} u_b) = \sum_{j=1}^J (u'_b \mathbf{Y}_j^{+b})^2 \geq 0$,

constata-se que $(\sum_{j=1}^J \mathbf{Y}_j^{+b})$ é semi-definida positiva.

Como $(\sum_{j=1}^J \mathbf{Y}_j^{+b})$ não é matriz nula, o seu maior valor próprio λ , a que estará associado o primeiro vector próprio $\nu_{1,b}$ será positivo (ver Pollock, 1979).

Agora $\nu_{1,b}$ terá de ser ortogonal a $\frac{1}{b^{1/2}} \mathbf{1}_b$, tendo soma das componentes nula (ver Giraldes et al., 1995). A demonstração fica completa observando que $\|\nu_{1,b}\| = 1$, e dado que $\nu_{1,b}$ satisfaz as condições $\sum_{i=1}^b p_i x_i = 0$ e $\sum_{i=1}^b p_i x_i^2 = 1$ e maximiza $M(\mathbf{x}_b)$ na família de todos os vectores, que satisfazem a segunda destas condições. ■

3.5.2 Estimadores de máxima verosimilhança

Dentro dos vários procedimentos tendo como objectivo a obtenção dos melhores estimadores, destacam-se o método dos mínimos quadrados, o método dos momentos e o método da máxima verosimilhança. Em particular este último tem especial importância na inferência estatística.

Admitamos que as produções são realizações de variáveis normais, independentes, com variâncias inversamente proporcionais aos pesos e sejam, $N = bJ$ o número de observações, x_1, \dots, x_b o vector dos índices ambientais e σ^2 um factor de proporcionalidade das variâncias.

Teremos a função de verosimilhança

$$L(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J, \mathbf{x}_b, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j \in D_i} p_i (y_{i,j} - \alpha_j - \beta_j x_i)^2}}{(2\pi)^{N/2} \sigma^N}$$

e logaritmizando vem

$$\ln(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J, \mathbf{x}_b, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j=1}^J p_i (y_{i,j} - \alpha_j - \beta_j x_i)^2 + C$$

com

$$C = -\frac{N}{2} (\ln(2\pi) + \ln(\sigma^2)).$$

Começemos por deduzir o estimador de σ^2 fazendo

$$\frac{\partial \ln(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J, \mathbf{x}_b, \sigma^2)}{\partial \sigma^2} = \frac{1}{2\sigma^4} \sum_{i=1}^b \sum_{j=1}^J p_i (y_{i,j} - \alpha_j - \beta_j x_i)^2 - \frac{N}{2\sigma^2}$$

e igualando a zero virá

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^b \sum_{j=1}^J p_i (y_{i,j} - \hat{\alpha}_j - \hat{\beta}_j \hat{x}_i)^2$$

Agora para obter $(\hat{\alpha}_j, \hat{\beta}_j)$, $j = 1, \dots, J$ e \hat{x}_i , $i = 1, \dots, b$ teremos que minimizar

$$\sum_{i=1}^b \sum_{j=1}^J p_i (y_{i,j} - \alpha_j - \beta_j x_i)^2.$$

Sendo assumida a normalidade os estimadores de mínimos quadrados são também estimadores de máxima verosimilhança.

3.5.3 Testes de razão de verosimilhanças

No caso completo temos, o número de observações $N = bJ$ e os pesos iguais a p_i , $j = 1, \dots, J$ e $i = 1, \dots, b$. No que se segue, restringir-nos-emos a regressões lineares, com pares de coeficientes (α_j, β_j) , $j = 1, \dots, J$.

Começaremos por testar a hipótese

$$H_0 : \beta_1 = \dots = \beta_J = \beta$$

sendo β o valor comum a estes coeficientes quando H_0 se verifica. Para obtermos o supremo da verosimilhança associado a H_0 teremos que maximizar

$$\ln(\alpha_J, \beta \mathbf{1}_J, x_b, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j=1}^J p_i(y_{i,j} - \alpha_j - \beta x_i)^2 - \frac{N}{2} (\ln(2\pi) + \ln(\sigma^2))$$

calculando

$$\frac{\partial \ln(\alpha_J, \beta \mathbf{1}_J, x_b, \sigma^2)}{\partial \sigma^2} = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^b \sum_{j=1}^J p_i(y_{i,j} - \alpha_j - \beta x_i)^2 - \frac{N}{2\sigma^2}$$

e igualando a zero, teremos

$$\hat{\sigma}_w^2 = \frac{1}{N} \sum_{i=1}^b \sum_{j=1}^J p_i(y_{i,j} - \hat{\alpha}_{jw} - \hat{\beta}_w \hat{x}_{iw})^2.$$

Para se obter $(\hat{\alpha}_j, \hat{\beta}_j)$, $j = 1, \dots, J$ e \hat{x}_i , $i = 1, \dots, b$ vamos minimizar $S_w(\alpha_J, \beta, \mathbf{x}_b) = \sum_{i=1}^b \sum_{j=1}^J p_i(y_{i,j} - \hat{\alpha}_{jw} - \hat{\beta}_w \hat{x}_{iw})^2$. Aplicando o algoritmo zig-zag, minimiza-se em cada iteração, primeiro em ordem a (α_j, β_j) e em seguida em ordem a \mathbf{x}_b .

Como valores iniciais podemos considerar $x_b^1 = \hat{x}_b$, ou seja, os valores dos estimadores de máxima verosimilhança sem restrições para os índices ambientais.

Derivando a função S_w em ordem aos coeficientes obtemos

$$\begin{cases} \frac{\partial S_w}{\partial \alpha_j} = -2 \sum_{i=1}^b p_i(y_{i,j} - \alpha_{j,w} - \beta_w x_{i,w}) & ; \quad j = 1, \dots, J \\ \frac{\partial S_w}{\partial \beta} = -2 \sum_{i=1}^b p_i x_{i,w} (y_{i,j} - \alpha_{j,w} - \beta_w x_{i,w}) \end{cases}$$

Igualando a zero as primeiras derivadas $\frac{\partial S_w}{\partial \alpha_j}, j = 1, \dots, J$, vem

$$-2 \sum_{i=1}^b p_i(y_{i,j} - \alpha_{j,w} - \beta_w x_{i,w}) = 0$$

pelo que

$$\beta_w = \frac{\sum_{i=1}^b p_i y_{i,j} - \sum_{i=1}^b p_i \alpha_{j,w}}{\sum_{i=1}^b p_i x_{i,w}}, \quad \text{com} \quad \alpha_{j,w} = \alpha_1 + d_{j,w}, \quad j = 2, \dots, J$$

sendo

$$d_{jw} = \frac{\frac{\sum_{i=1}^b p_i y_{i,j}}{b} - \frac{\sum_{i=1}^b p_i y_{i,1}}{b}}{\frac{\sum_{i=1}^b p_i x_{i,w}}{b}} = \frac{\sum_{i=1}^b p_i (y_{i,j} - y_{i,1})}{\sum_{i=1}^b p_i x_{i,w}}, \quad j = 2, \dots, J$$

Analogamente igualando a zero as derivadas $\frac{\partial S_w}{\partial \beta}$, vem

$$-2 \sum_{i=1}^b \sum_{j=1}^J p_i x_{i,w} (y_{i,j} - \alpha_{j,w} - \beta_w x_{i,w}) = 0$$

pelo que virá

$$\beta_w = \frac{\sum_{i=1}^b \sum_{j=1}^J p_i x_{i,w} y_{i,j} - \sum_{i=1}^b (\sum_{j=1}^J p_i x_{i,w}) (\alpha_1 + d_{j,w})}{\sum_{i=1}^b p_i x_{i,w}^2}$$

De ambas as expressões para β_w , conclui-se que $\alpha_1 = \frac{A_1 - A_2}{B_1 - B_2}$, com

$$\left\{ \begin{array}{l} A_1 = \frac{\sum_{i=1}^b p_i \sum_{j=1}^J y_{i,j} - \sum_{i=1}^b p_i \sum_{j=1}^J d_{j,w}}{J \sum_{i=1}^b p_i x_{i,w}} \\ A_2 = \frac{\sum_{i=1}^b \sum_{j=1}^J p_i x_{i,w} y_{i,j} - \sum_{i=1}^b p_i x_{i,w} \sum_{j=1}^J d_{j,w}}{J \sum_{i=1}^b p_i x_{i,w}^2} \\ B_1 = \frac{\sum_{i=1}^b p_i}{\sum_{i=1}^b p_i x_{i,w}} \\ B_2 = \frac{\sum_{i=1}^b p_i x_{i,w}}{\sum_{i=1}^b p_i x_{i,w}^2} \end{array} \right.$$

A partir daqui obtêm-se os $\alpha_2, \dots, \alpha_J$ e o β correspondente à primeira iteração passando-se de seguida à maximização em ordem aos índices ambientais. A função a maximizar é

$$\sum_{i=1}^b h_i(x_i) = \sum_{i=1}^b \sum_{j=1}^J p_i (y_{i,j} - \alpha_j - \beta x_i)^2, \quad i = 1, \dots, b.$$

Derivando em ordem a x_i e igualando a zero vem

$$\sum_{j=1}^J p_i (y_{i,j} - \alpha_j - \beta x_i) = 0$$

e logo

$$x_i = \frac{\sum_{j=1}^J (y_{i,j} - \alpha_j)}{J\beta}, \quad i = 1, \dots, b.$$

Passemos de seguida ao teste

$$H_0 : \alpha_1 = \dots = \alpha_J = \alpha$$

onde α representa o valor comum das ordenadas na origem, quando H_0 se verifica, e w o espaço paramétrico correspondente.

Neste caso teremos

$$\hat{\sigma}_w^2 = \frac{1}{bJ} \sum_{i=1}^b p_i \sum_{j=1}^J (y_{i,j} - \hat{\alpha}_w - \hat{\beta}_{j,w} \hat{x}_{i,w})^2$$

o que nos leva a minimizar através do algoritmo zig-zag a função

$$S_w(\alpha, \beta_J, x_b) = \sum_{i=1}^b \sum_{j=1}^J p_i (y_{i,j} - \alpha - \beta_{j,w} x_{i,w})^2$$

para obter $\hat{\alpha}_w$, $\hat{\beta}_w$ e $\hat{x}_{w,b}$.

Será $\beta_{j,w} = \beta_1 + d_{j,w}$, $j = 2, \dots, J$,

com

$$d_{j,w} = \frac{\frac{\sum_{i=1}^b p_i y_{i,j} x_{i,w}}{b} - \frac{\sum_{i=1}^b p_i y_{i,1} x_{i,w}}{b}}{\frac{\sum_{i=1}^b p_i x_{i,w}}{b} - \frac{\sum_{i=1}^b p_i x_{i,w}}{b}} = \frac{\sum_{i=1}^b p_i x_{i,w} (y_{i,j} - y_{i,1})}{\sum_{i=1}^b p_i x_{i,w}^2 - \sum_{i=1}^b p_i x_{i,w}}, \quad j = 2, \dots, J$$

vindo

$$\left\{ \begin{array}{l} A_1 = \frac{\sum_{i=1}^b p_i x_{i,w} \sum_{j=1}^J y_{i,j} - \sum_{i=1}^b p_i x_{i,w}^2 \sum_{j=1}^J d_{j,w}}{J \sum_{i=1}^b p_i x_{i,w}} \\ A_2 = \frac{\sum_{i=1}^b p_i \sum_{j=1}^J y_{i,j} - \sum_{i=1}^b p_i x_{i,w} \sum_{j=1}^J d_{j,w}}{J \sum_{i=1}^b p_i} \\ B_1 = \frac{\sum_{i=1}^b p_i x_{i,w}^2}{b} \\ B_2 = \frac{\sum_{i=1}^b p_i x_{i,w}}{b} \end{array} \right.$$

Na segunda parte, ou seja na maximização em ordem aos índices ambientais pode tomar-se

$$x_i = \frac{\sum_{j=1}^J (y_{i,j} - \alpha) \beta_j}{\sum_{j=1}^J \beta_j^2}, \quad i = 1, \dots, b$$

Por último vamos testar

$$H_0 : \begin{cases} \alpha_1 = \dots = \alpha_J = \alpha \\ \beta_1 = \dots = \beta_J = \beta \end{cases}$$

Neste caso

$$\hat{\sigma}_w^2 = \frac{1}{bJ} \sum_{i=1}^b p_i \sum_{j=1}^J (y_{i,j} - \hat{\alpha}_w - \hat{\beta}_{j,w} \hat{x}_{i,w})^2$$

e vamos minimizar

$$S_w(\alpha, \beta, \mathbf{x}_b) = \sum_{i=1}^b p_i \sum_{j=1}^J (y_{i,j} - \alpha - \beta x_{i,w})^2$$

Derivando agora a função S_w em ordem aos coeficientes, obtemos

$$\begin{cases} \frac{\partial S_w}{\partial \alpha} = -2 \sum_{i=1}^b p_i (y_{i,j} - \alpha - \beta x_{i,w}) & ; \quad j = 1, \dots, J \\ \frac{\partial S_w}{\partial \beta} = -2 \sum_{i=1}^b \sum_{j=1}^J p_i x_{i,w} (y_{i,j} - \alpha - \beta x_{i,w}) \end{cases}$$

e igualando as derivadas a zero obtemos um sistema do tipo

$$\begin{cases} a_{11}\alpha + a_{12}\beta = b_1 \\ a_{21}\alpha + a_{22}\beta = b_2 \end{cases}$$

cuja solução será

$$\begin{cases} \alpha = \frac{a_{21}b_1 - a_{12}b_2}{d} \\ \beta = \frac{a_{11}b_2 - a_{21}b_1}{d} \end{cases}$$

Neste caso, tem-se $a_{11} = J \sum_{i=1}^b p_i$; $a_{12} = a_{21} = J \sum_{i=1}^b p_i x_{i,w}$;

$$a_{22} = J \sum_{i=1}^b p_i x_{i,w}^2 \quad ; \quad b_1 = \sum_{i=1}^b p_i \sum_{j=1}^J y_{i,j} \quad e \quad b_2 = \sum_{i=1}^b p_i x_{i,w} \sum_{j=1}^J y_{i,j},$$

vindo

$$d = J^2 \left[\left(\sum_{i=1}^b p_i \right) \left(\sum_{i=1}^b p_i x_{i,w}^2 \right) - \left(\sum_{i=1}^b p_i x_{i,w} \right)^2 \right]$$

Na segunda parte da iteração obtém-se directamente

$$x_i = \frac{1}{J\beta} \sum_{j=1}^J (y_{i,j} - \alpha) \quad ; \quad i = 1, \dots, b$$

3.6 Redes Fraccionalmente Completas

Consideremos ensaios de comparação de cultivares integrados em redes de ensaios que podem ser decompostas em sub-redes conectadas. O plano será constituído por várias sub-redes, cada uma com os respectivos conjuntos de locais e de cultivares utilizados definidos. A título ilustrativo uma rede de ensaios pode ser constituída por observações de vários anos e cada ano pode constituir uma sub-rede (cada ano terá informação dos locais e cultivares utilizados).

Definição: *Dada uma sub-rede dizemos que estamos perante o caso completo, quando o conjunto dos cultivares afecto a essa sub-rede é o mesmo para todos os locais da sub-rede.*

Definição: *Uma rede diz-se fraccionalmente completa se cada sub-rede que a constitui for completa.*

Uma rede completa terá o mesmo conjunto de cultivares em todos os locais e portanto uma rede pode ser fraccionalmente completa sem ser completa, uma vez que o conjunto de cultivares pode variar de sub-rede para sub-rede.

Considere-se uma rede global e sejam as sub-redes que a constituem numeradas de 1 a L. Temos assim

Proposição 3.5 *Se sub-redes contíguas numa rede fraccionalmente completa tiverem cultivares em comum, a rede global diz-se conectada.*

Dem: Dados os blocos b_i $[b'_i]$ pertencentes à sub-rede respectiva r_i $[r'_i]$, para construir uma cadeia do primeiro para o segundo bloco é possível construir uma cadeia que passe sucessivamente em todas as cadeias intermédias. Sendo as sub-redes 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 pelo menos um cultivar comum. ■

No caso das redes fraccionalmente completas conectadas, existe conexão na rede global e pode proceder-se a um ajustamento conjunto em simultâneo com o ajustamento sub-rede a sub-rede.

Considere-se as funções objectivo das sub-redes de uma rede fraccionalmente completa, S_1, \dots, S_L e seja S a correspondente função objectivo da rede global.

Denote-se por S^* o mínimo da função objectivo global e por S_1^*, \dots, S_L^* os mínimos das funções objectivo parciais. Como visto em Pinto(2006), o mínimo da função objectivo global é superior à soma dos mínimos das funções objectivo parciais, ou seja

$$S^* > \sum_{\ell=1}^L S_\ell^* \quad (3.6.23)$$

3.7 Método do Contorno Superior

A utilização do método do contorno superior na comparação e selecção de cultivares foi introduzida por Mexia *at al.* (1997). Na abordagem a este assunto seguiremos de perto esse trabalho assim como os de Pereira (2004) e Pinto (2006) onde o mesmo tema é desenvolvido. Começemos então por considerar j regressões ajustadas

$$\tilde{y}_j = \tilde{\alpha}_j + \tilde{\beta}_j x ; j = 1, \dots, J$$

uma por cultivar e seja $[\theta_{min}; \theta_{max}]$ o intervalo cujos extremos são exactamente o mínimo e o máximo dos índices ambientais ajustados. De entre todas as regressões, teremos no intervalo $[\theta_{min}; \theta_{max}]$ as regressões dominantes que são aquelas que correspondem às produções máximas, logo integrando o contorno superior. Os cultivares associados às regressões dominantes serão naturalmente os cultivares dominantes. Na selecção dos cultivares a recomendar, deve obviamente ser dada principal atenção aos que correspondem às regressões que integram o contorno superior. Estes serão comparados com todos os restantes, e no caso de algum destes ser significativamente dominado, deverá ser eliminado do processo de selecção.

3.7.1 Geometria do contorno superior

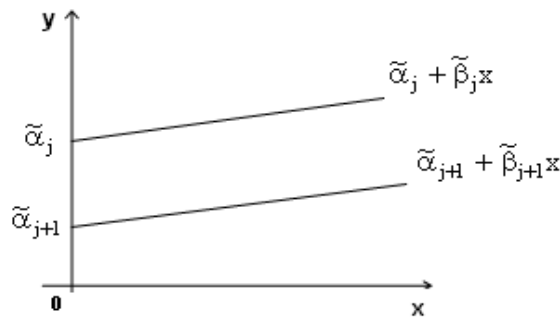
Sendo as regressões ajustadas do tipo linear, o contorno superior será uma poligonal aberta composta por segmentos de recta.

Atribuindo índices j com $j = 1, \dots, J$, a essas regressões ajustadas, e ordenando-as por ordem decrescente dos seus declives ficará

$$\tilde{\beta}_1 \geq \tilde{\beta}_2 \geq \tilde{\beta}_J$$

tendo-se $\tilde{\alpha}_j > \tilde{\alpha}_{j+1}$ no caso particular em que se tenha $\tilde{\beta}_j = \tilde{\beta}_{j+1}$. No caso de termos rectas paralelas apenas a que tiver maior ordenada na origem fará parte do contorno superior e consequentemente apenas esse cultivar poderá ser dominante, como se pode observar na figura seguinte com o exemplo de duas rectas paralelas

Figura 3.1: Rectas paralelas



Vamos prosseguir a análise deste assunto estabelecendo a

Proposição 3.6 *O contorno superior é uma poligonal aberta convexa.*

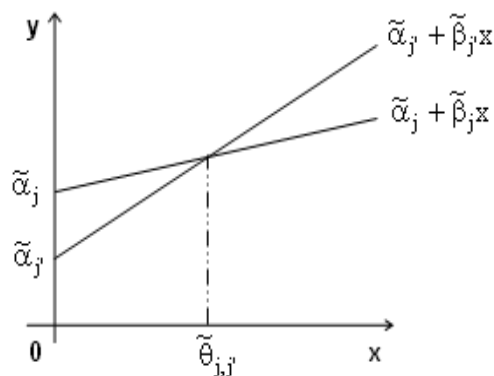
Dem: O contorno superior é uma poligonal aberta constituída por segmentos de recta pertencentes às regressões dominantes. Quando duas rectas j e j' se intersectam no ponto de abcissa $\theta_{j,j'}$ e temos a relação entre declives $\tilde{\beta}_1 > \tilde{\beta}_2$, então à direita desse ponto verifica-se $\tilde{\alpha}_j + \tilde{\beta}_j x > \tilde{\alpha}_{j'} + \tilde{\beta}_{j'} x$. Deste modo, temos as inclinações dos segmentos de recta que formam o contorno superior a crescer da esquerda para a direita o que completa a demonstração. ■

Quando duas rectas têm declives diferentes, ou seja $\tilde{\beta}_j \neq \tilde{\beta}_{j'}$, temos

$$\tilde{\theta}_{j,j'} = \frac{\tilde{\alpha}_j - \tilde{\alpha}_{j'}}{\tilde{\beta}_{j'} - \tilde{\beta}_j}, j \neq j' \text{ e } j, j' = 1, \dots, J \quad (3.7.24)$$

e o ponto de intersecção terá as coordenadas $(\tilde{\theta}_{j,j'}, \tilde{\alpha}_j + \tilde{\beta}_j \tilde{\theta}_{j,j'})$. Na figura 3.2 ilustramos o caso que acabamos de descrever com $j < j'$.

Figura 3.2: Intersecção das rectas ajustadas de índices j e j'



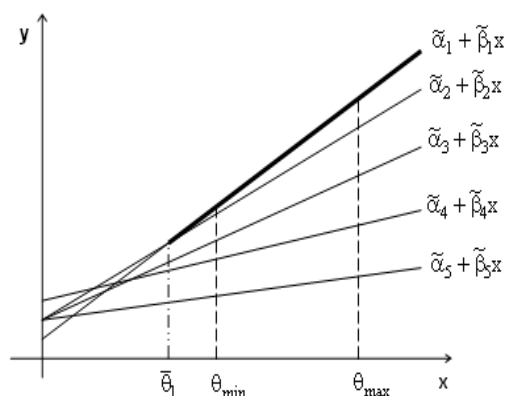
Para a determinação da poligonal que forma o contorno superior, apresenta-se a discussão de três das situações mais comuns, exemplificando-se com cinco cultivares. Como ponto de partida sugere-se o seguinte procedimento

- i) Determinar $\bar{\theta}_1 = \max\{\tilde{\theta}_{1,j}, j = 2, \dots, J\}$
- ii) Localizar $\bar{\theta}_1$ no intervalo $[\theta_{min}, \theta_{max}]$

e de seguida estaremos perante um dos seguintes casos:

1º Caso - $\bar{\theta}_1 < \theta_{min}$

Nesta situação o contorno superior corresponde exactamente à primeira recta de regressão ajustada $\tilde{\alpha}_1 + \tilde{\beta}_1 x$ e apenas o primeiro cultivar será dominante conforme se ilustra na figura 3.3, para $J = 5$.

Figura 3.3: Traçado do contorno superior para $\bar{\theta}_1 \leq \theta_{min}$ 

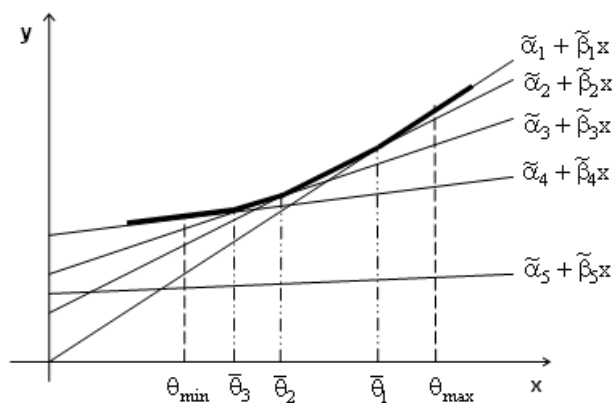
2º Caso - $\theta_{min} < \bar{\theta}_1 < \theta_{max}$

Neste caso, no intervalo $[\bar{\theta}_1, \theta_{max}]$, ter-se-á $\tilde{\alpha}_j + \tilde{\beta}_j x < \tilde{\alpha}_1 + \tilde{\beta}_1 x$, $j = 2, \dots, J$, pelo que o cultivar 1 é dominante. Haverá para além do primeiro mais cultivares dominantes no intervalo $[\theta_{min}, \bar{\theta}_1]$. Neste intervalo o cultivar com índice j é dominante e para se averiguar da existência de mais cultivares dominantes terá de se obter

$$\bar{\theta}_2 = \max\{\tilde{\theta}_{j,j'}, j' = j + 1, \dots, J\}$$

Para os intervalos seguintes terá de se proceder de forma análoga até que o contorno superior esteja completamente definido.

Na figura 3.4, apresentamos um exemplo para $J = 5$, onde se verifica a existência de quatro cultivares dominantes.

Figura 3.4: Traçado do contorno superior para $\theta_{min} < \bar{\theta}_1 \leq \theta_{max}$ 

3º Caso - $\bar{\theta}_1 > \theta_{max}$

Neste caso a recta ajustada com índice 1 não faz parte do contorno superior, pelo que, o cultivar correspondente não é dominante. Teremos assim que averiguar relativamente ao conjunto de rectas com índices $j = 2, \dots, J$ determinado

i) $\bar{\theta}_2 = \max\{\theta_2, j = 3, \dots, J\}$

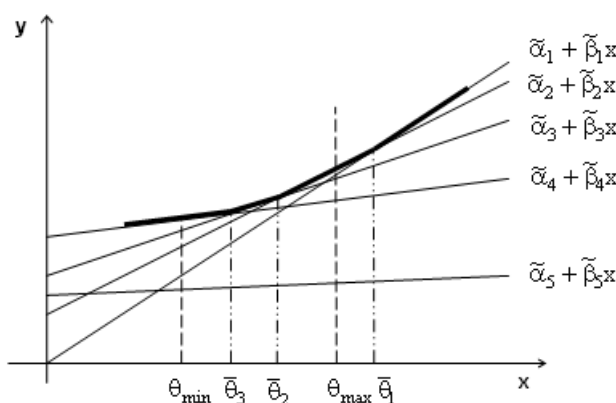
ii) se $\bar{\theta}_2 \geq \theta_{max}$ o cultivar com índice 2 não é dominante e passamos a analisar as rectas sobrantes

iii) se $\bar{\theta}_2 \in [\theta_{min}, \theta_{max}]$ deve seguir-se o procedimento apresentado no 2º Caso

iv) continua-se o processo até que todo o contorno superior esteja definido.

Na figura 3.5, apresenta-se um exemplo para $J = 5$, onde se verifica a existência de três cultivares dominantes.

Figura 3.5: Traçado do contorno superior para $\bar{\theta}_1 \geq \theta_{max}$



Tendo em atenção as diferentes possibilidades que podem ocorrer no 2º Caso, interessa averiguar com mais detalhe alguns aspectos que se revelam importantes. Estabeleçamos então a

Proposição 3.7 Se $\bar{\theta}_1 = \tilde{\theta}_{1,j'}$ e para $j \neq 1, j', \tilde{\theta}_{1,j'} > \tilde{\theta}_{1,j}$ o cultivar j' é dominante.

Dem: Tendo-se $j \neq 1, j'$ e $\tilde{\theta}_{1,j} < x < \tilde{\theta}_{1,j'}$ verificam-se as desigualdades $\tilde{\alpha}_j + \tilde{\beta}_j x < \tilde{\alpha}_1 + \tilde{\beta}_1 x < \tilde{\alpha}_{j'} + \tilde{\beta}_{j'} x$ e portanto como deste modo o cultivar j' é dominante verifica-se a tese. ■

Suponhamos agora que no ponto $(\bar{\theta}_1, \tilde{\alpha}_1 + \tilde{\beta}_1\bar{\theta}_1)$ a recta de índice 1 é intersectada por mais do que uma recta de regressão. Seja Λ o conjunto dos índices dessas rectas de regressão ajustadas e j' o maior dos índices desse conjunto Λ . Estabeleçamos a proposição seguinte

Proposição 3.8 *Se j' é o maior índice para o qual $\bar{\theta}_1 = \tilde{\theta}_{1,j'}$ o cultivar com índice j' é dominante.*

Dem: A proposição estabelece que das rectas que intersectam a recta $\tilde{\alpha}_1 + \tilde{\beta}_1x$, em $\bar{\theta}_1$ pertencerá ao contorno superior à esquerda de $\bar{\theta}_1$ aquela que tiver menor inclinação. Além disso, se $1 < j < j'$ então $\tilde{\beta}_{j'} < \tilde{\beta}_j$, $j = 2, \dots, j' - 1$, pois caso contrário ter-se-ia $(\tilde{\alpha}_{j'} < \tilde{\alpha}_j) \wedge (\tilde{\theta}_{1,j'} < \tilde{\theta}_{1,j})$ o que é impossível dado que neste caso $\bar{\theta}_1 = \tilde{\theta}_{1,j'}$. Se Λ for o conjunto dos índices das regressões ajustadas que intersectam a recta $\tilde{\alpha}_1 + \tilde{\beta}_1x$ em $\bar{\theta}_1$, o maior índice em Λ é j' . Se j é outro índice de Λ , tem-se $\tilde{\beta}_{j'} < \tilde{\beta}_j$, pelo que $\tilde{\alpha}_j + \tilde{\beta}_jx < \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}x$ para $x < \bar{\theta}_1 \wedge j \in \Lambda$. Finalmente, se $j \notin \Lambda$, temos $\tilde{\theta}_{1,j} < \tilde{\theta}_{1,j'} = \bar{\theta}_1$ e sempre que $\tilde{\theta}_{1,j} < x < \tilde{\theta}_{1,j'} = \bar{\theta}_1$, verifica-se $\tilde{\alpha}_{j'} + \tilde{\beta}_{j'}x > \tilde{\alpha}_1 + \tilde{\beta}_1x > \tilde{\alpha}_j + \tilde{\beta}_jx$. Deste modo constata-se a existência à esquerda de θ_1 dum intervalo de dominância para o cultivar com índice j . ■

Proposição 3.9 *Sendo $1 < j < j'$, com j' o maior índice no conjunto Λ , o cultivar com índice j não é dominante.*

Dem: Tem-se $\tilde{\beta}_1 > \tilde{\beta}_j > \tilde{\beta}_{j'}$ e também $\tilde{\alpha}_j + \tilde{\beta}_j\bar{\theta}_1 \leq \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\bar{\theta}_1 = \tilde{\alpha}_1 + \tilde{\beta}_1\bar{\theta}_1$, pelo que quando $x < \bar{\theta}_1$ ter-se-á $\tilde{\alpha}_j + \tilde{\beta}_jx < \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}x$ e no caso de $x > \bar{\theta}_1$ ter-se-á $\tilde{\alpha}_j + \tilde{\beta}_jx < \tilde{\alpha}_1 + \tilde{\beta}_1x$, verificando-se de imediato que em nenhum dos casos o cultivar com índice j é dominante, o que estabelece a tese. ■

3.7.2 Comparação e selecção de cultivares

Seguindo Pereira(2004) onde na abordagem a este assunto, se admite a utilização de índices ambientais L_2 e assumindo que

$$[\tilde{\alpha}_j, \tilde{\beta}_j]' \sim N([\alpha_j, \beta_j]'; \sigma^2 \mathbf{W}_j), j = 1, \dots, J \quad (3.7.25)$$

sendo $\mathbf{W}_j = (\mathbf{X}_j' \mathbf{D}_j \mathbf{X}_j)^{-1}$, $j = 1, \dots, J$ independente de $S \sim \sigma^2 \chi_g^2$ e com $g = \sum_{j=1}^J b_j - 2J$.

Vimos que sendo j' o índice de um dos cultivares dominantes e $[\theta_{min}^{j'}, \theta_{max}^{j'}]$ o respectivo intervalo de dominância, há que considerar os seguintes casos

Caso geral

Teremos que analisar as hipóteses

$$H_0^{j,j'}(\theta_0) : U_j(\theta_0) = U_{j'}(\theta_0) \quad (3.7.26)$$

versus

$$H_1^{j,j'}(\theta_0) : U_j(\theta_0) < U_{j'}(\theta_0) \quad (3.7.27)$$

sendo

$$U_j(\theta_0) = \alpha_j + \beta_j \theta_0, j = 1, \dots, J \quad (3.7.28)$$

$$U_{j'}(\theta_0) = \alpha_{j'} + \beta_{j'} \theta_0, j' = 1, \dots, J \quad (3.7.29)$$

e $j \neq j'$.

Para testar as hipóteses consideradas deverá ser usada a estatística de teste

$$t_{j,j'}(\theta_0) = \frac{\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)}{\sqrt{\frac{S}{g}[k_j(\theta_0) + k_{j'}(\theta_0)]}} \quad (3.7.30)$$

que quando H_0 se verifica tem distribuição t central, com g graus de liberdade e sendo $k_j(\theta_0) = [1:\theta_0]W_j[1:\theta_0]'$, $j = 1, \dots, J$.

1º Caso particular : $j' < j$

Para estes casos teremos $\beta_{j'} > \beta_j$ significando que no intervalo $[\theta_{min}^{j'}, \theta_{max}^{j'}]$ as rectas estão mais próximas quando $x = \theta_{min}^{j'}$. Assim, será de testar a hipótese

$$H_0^{j,j'} : \alpha_j + \beta_j \theta_{min}^{j'} = \alpha_{j'} + \beta_{j'} \theta_{min}^{j'}$$

versus

$$H_1^{j,j'} : \alpha_j + \beta_j \theta_{min}^{j'} < \alpha_{j'} + \beta_{j'} \theta_{min}^{j'}$$

2º Caso particular : $j' > j$

Aqui temos $\beta_j > \beta_{j'}$ e portanto neste caso e considerando o mesmo intervalo de dominância $[\theta_{min}^{j'}, \theta_{max}^{j'}]$ as rectas estão mais próximas quando $x = \theta_{max}^{j'}$.

Na comparação dos cultivares j' e j somos então levados a testar a hipótese

$$H_0^{j,j'} : \alpha_j + \beta_j \theta_{max}^{j'} = \alpha_{j'} + \beta_{j'} \theta_{max}^{j'}$$

versus

$$H_1^{j,j'} : \alpha_j + \beta_j \theta_{max}^{j'} < \alpha_{j'} + \beta_{j'} \theta_{max}^{j'}$$

No caso do primeiro cultivar dominante temos $\theta_{max}^{j'} = \theta_{max}$ devendo considerar-se apenas o primeiro par de hipóteses, enquanto que no caso do último cultivar dominante será $\theta_{min}^{j'} = \theta_{min}$ e deve considerar-se apenas o segundo par de hipóteses.

Por forma a obter-se uma maior robustez podemos aplicar os métodos de comparação múltipla de Scheffé, de Bonferroni e de Tuckey, que passamos a descrever.

Método de comparação múltipla de Scheffé

Seguindo Scheffé(1959, págs 66 a 70), considerando $f_{1-\alpha, g_1, g}$ o valor crítico da distribuição F de Snedecor ao nível de significância α , com g_1 e g graus de liberdade, teremos

i) $U_1(\theta_0), \dots, U_{j'-1}(\theta_0)$ que verifiquem

$$\frac{|\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)|}{\sqrt{\frac{\Sigma}{g}(j'-1)(k_j(\theta_0) + k_{j'}(\theta_0))f_{1-\alpha, j'-1, g}}} > 1, \quad j = 1, \dots, j' - 1$$

são significativamente inferiores a $U_{j'}(\theta_0)$ ao nível de significância α , e por outro lado

ii) $U_{j'+1}(\theta_0), \dots, U_J(\theta_0)$ que verifiquem

$$\frac{|\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)|}{\sqrt{\frac{\Sigma}{g}(J-j')(k_j(\theta_0) + k_{j'}(\theta_0))f_{1-\alpha, J-j', g}}} > 1, \quad j = j' + 1, \dots, J$$

são significativamente inferiores a $U_{j'}(\theta_0)$ ao nível de significância α .

Método de comparação múltipla de Bonferroni

Em certos casos em que há possibilidade de melhorar os resultados obtidos pelo método de comparação múltipla de Scheffé, podemos usar a desigualdade de Bonferroni. Quando o objectivo é a comparação de pares de observações podemos substituir nas expressões anteriores, ver (Seber, 2003) as quantidades $\sqrt{(j'-1)f_{1-\alpha, j'-1, g}}$ e $\sqrt{(J-j')f_{1-\alpha, J-j', g}}$ pelos quantis da distribuição t de Student com g graus de liberdade, nos níveis de significância $\frac{\alpha}{2(j'-1)}$ e $\frac{\alpha}{2(J-j')}$ respectivamente.

Então teremos

i) $U_1(\theta_0), \dots, U_{j'-1}(\theta_0)$ que verifiquem

$$\frac{|\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)|}{\sqrt{\frac{\tilde{S}}{g}(k_j(\theta_0) + k_{j'}(\theta_0)) t_{1-\frac{\alpha}{2(j'-1)}, g}}} > 1, \quad j = 1, \dots, j' - 1$$

são significativamente inferiores a $U_{j'}(\theta_0)$ ao nível de significância $\frac{\alpha}{2(j'-1)}$, e por outro lado

ii) $U_{j'+1}(\theta_0), \dots, U_J(\theta_0)$ que verifiquem

$$\frac{|\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)|}{\sqrt{\frac{\tilde{S}}{g}(k_j(\theta_0) + k_{j'}(\theta_0)) t_{1-\frac{\alpha}{2(J-j')}, g}}} > 1, \quad j = j' + 1, \dots, J$$

são significativamente inferiores a $U_{j'}(\theta_0)$ ao nível de significância $\frac{\alpha}{2(J-j')}$.

Método de comparação múltipla de Tukey

No caso completo, ver Scheffé(1959) é possível aplicar o método de Tukey. Neste caso os $U_1(\theta_0), \dots, U_{j'-1}(\theta_0)$ para os quais se verifica a condição

$$\frac{|\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)|}{\sqrt{\frac{\tilde{S}}{g} k(\theta_0) (t_{1-\alpha, j', g})^2}} > 1, \quad j = 1, \dots, j' - 1$$

onde $t_{1-\alpha, g_1, g}$ representa o quantil da distribuição t de Student com g_1 e g graus de liberdade ao nível de significância α , são conjunta e significativamente inferiores a $U_{j'}(\theta_0)$ ao nível de significância α .

Por outro lado os $U_{j'+1}(\theta_0), \dots, U_J(\theta_0)$ que verifiquem

$$\frac{|\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)|}{\sqrt{\frac{\tilde{S}}{g} k(\theta_0) (t_{1-\alpha, J-j'+1, g})^2}} > 1, \quad j = j' + 1, \dots, J$$

são conjunta e significativamente inferiores a $U_{j'}(\theta_0)$ ao nível de significância α . Uma vez considerado o caso completo teremos

$$k(\theta_0) = [1 \quad \theta_0] \mathbf{W} [1 \quad \theta_0]'$$

onde $\mathbf{W} = (\mathbf{X}' \mathbf{D} \mathbf{X})^{-1}$, sendo $\mathbf{X} = [1^n \quad \mathbf{x}^n]'$ e \mathbf{D} a matriz diagonal dos pesos dos blocos.

3.8 Estabilidade dos Cultivares

A existência de estabilidade no comportamento dos cultivares é traduzida pela ausência de diferenças significativas nos coeficientes angulares das regressões ajustadas. Esta característica tem grande importância na medida em que do ponto de vista da condução de um plano de melhoramento é valorizada a obtenção de cultivares que tenham boa estabilidade de resposta às variações dos índices ambientais.

Para verificação de existência de estabilidade dos cultivares ao longo dos anos apresenta-se o seguinte teste

$$H_{0,j} : \beta_{j,i_1} = \dots = \beta_{j,i_j}, j = 1, \dots, J \quad (3.8.31)$$

onde $\{i_1, \dots, i_j\}$ representa o conjunto dos índices relativos aos anos em que o j -ésimo cultivar foi utilizado. Por forma a chegarmos à estatística do teste começemos por estabelecer

Lema 3.3 *Seja \mathbf{K} uma matriz definida positiva e \mathbf{A} uma matriz que se obtém quando é retirada a primeira linha igual a $\frac{1}{\sqrt{m}}\mathbf{1}_m$ a uma matriz ortogonal. As matrizes \mathbf{AKA}' e $(\mathbf{AKA}')^{-1}$ são definidas positivas.*

Dem: Sendo \mathbf{K} uma matriz definida positiva, podemos escrever $\mathbf{K} = \mathbf{P}'\mathbf{D}(r_1, \dots, r_m)\mathbf{P}$ com \mathbf{P} matriz ortogonal e $r_i > 0, i = 1, \dots, m$ os valores próprios positivos de \mathbf{K} . A matriz $\mathbf{K}^{1/2} = \mathbf{P}'\mathbf{D}(r_1^{1/2}, \dots, r_m^{1/2})\mathbf{P}$ também é definida positiva e tem-se $\mathbf{K}^{1/2}\mathbf{K}^{1/2} = \mathbf{K}$. Deste modo temos $\mathbf{AKA}' = \mathbf{AK}^{1/2}\mathbf{A}' = (\mathbf{AK}^{1/2})(\mathbf{AK}^{1/2})'$. e como $\mathbf{K}^{1/2}$ é regular ter-se-á, ver Silvey(1975), $\text{Car}(\mathbf{AK}^{1/2}) = \text{car}(\mathbf{AKA}') = \text{car}(\mathbf{A}) = m - 1$ pelo que \mathbf{AKA}' será regular e sendo para $\forall w \neq 0, w'(\mathbf{AKA}')w = \|\mathbf{AK}^{1/2}w\|^2 > 0$, então \mathbf{AKA}' será definida positiva. A demonstração fica completa observando-se que se uma matriz é definida positiva a sua inversa também o é, ver Judge(1985). ■

Estabeleçamos agora

Proposição 3.10 *Seja $\underline{\mathbf{Z}}_m \sim N(\underline{\boldsymbol{\eta}}_m, \sigma^2\mathbf{K})$ com \mathbf{K} matriz definida positiva, independente de $S \sim \sigma^2\chi_g^2$ e \mathbf{A} a matriz obtida no Lema 3.3.*

A estatística

$$F = \frac{g}{r} \cdot \frac{(\mathbf{AZ}_m)'(\mathbf{AKA}')^{-1}\mathbf{AZ}_m}{S} \quad (3.8.32)$$

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

Dem: Em Seber (1980, pág. 5), podemos ver que $\mathbf{A}\underline{\mathbf{Z}}_m \sim N(\mathbf{A}\underline{\boldsymbol{\eta}}_m, \sigma^2 \mathbf{A}\mathbf{K}\mathbf{A}')$, pelo que, ver Mexia (1995, pág. 52), $(\mathbf{A}\underline{\mathbf{Z}}_m)'(\mathbf{A}\mathbf{K}\mathbf{A}')^{-1}\mathbf{A}\underline{\mathbf{Z}}_m \sim \sigma^2 \chi_{m-1, \delta}^2$ independente de $S \sim \sigma^2 \chi_g^2$, donde resulta a primeira parte da tese. Atendendo ao Lema 3.3, a matriz $\mathbf{A}\mathbf{K}\mathbf{A}'$ é definida positiva, pelo que podemos escrever, $\mathbf{A}\mathbf{K}\mathbf{A}' = \dot{\mathbf{P}}' \mathbf{D}(r_1, \dots, r_{m-1}) \dot{\mathbf{P}}$, com $\dot{\mathbf{P}}$ matriz ortogonal sendo $r_i > 0, i = 1, \dots, r$, os valores próprios de $\mathbf{A}\mathbf{K}\mathbf{A}'$. Ter-se-á então, $(\mathbf{A}\mathbf{K}\mathbf{A}')^{-1} = \dot{\mathbf{P}}' \mathbf{D}(r_1^{-1}, \dots, r_r^{-1}) \dot{\mathbf{P}}$, vindo $\delta = 0$ se e só se $\dot{\mathbf{P}} \mathbf{A} \underline{\boldsymbol{\eta}}_m = \mathbf{0}_r$, o que, por $\dot{\mathbf{P}}$ ser regular, equivale a ter-se $\mathbf{A} \underline{\boldsymbol{\eta}}_m = \mathbf{0}_r$. Ora, \mathbf{A} é constituída por r linhas duma matriz ortogonal de ordem m , tendo-se pois $\text{Car}(\mathbf{A}) = r$ 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 $\frac{c}{\sqrt{m}} \mathbf{1}_m$. Teremos então, $\delta = 0$ se e só se $\underline{\boldsymbol{\eta}}_m = \frac{c}{\sqrt{m}} \mathbf{1}_m$, o que estabelece a segunda parte da tese. ■

Na abordagem prática, iremos aplicar ao caso em que temos coeficientes ajustados em cada ano, pelo que sendo as somas dos quadrados dos desvios para a média dos índices ambientais ajustados nesse ano s_{a_l} , em que $a_l, l = 1, \dots, m_j$ é o índice correspondente, para o j -ésimo cultivar tomamos

$$\mathbf{K} = \mathbf{D}\left(\frac{1}{s_{a_1}}, \dots, \frac{1}{s_{a_{m_j}}}\right) \quad (3.8.33)$$

e

$$\underline{\mathbf{Z}}'_m = \left[\tilde{\beta}_{j, a_1}, \dots, \tilde{\beta}_{j, a_{m_j}} \right] \quad (3.8.34)$$

De notar ainda que S representa a soma das somas dos quadrados dos resíduos dos índices ambientais e o número de graus de liberdade $r = m - 1$.

Capítulo 4

Análise Conjunta de Regressões Lineares Completadas

4.1 Cultivares com Resposta Óptima

No presente capítulo, o principal objectivo é estabelecer a forma para seleccionar cultivares com capacidade óptima de resposta em índices ambientais altos. Para tal surge-nos um problema, onde não só se deve atender à produção física como também ao valor da mesma. A abordagem, passará pela consideração dos valores de outra variável, neste caso o peso específico, estimando-se coeficientes que traduzam o valor da produção dos vários cultivares. Deste modo, passa-se a aplicar a Análise Conjunta de Regressões às produções corrigidas dadas pelos produtos dos coeficientes estimados pelas produções físicas.

Interessará averiguar se o método clássico e o método corrigido conduzem à escolha do mesmo cultivar como tendo capacidade óptima de resposta. Na presente análise vamos aplicar dois métodos distintos

- o método do contorno superior aplicado às regressões ajustadas para as produções física e corrigida, correspondentes aos diferentes cultivares.
- o método da inclinação máxima que conduz à "selecção" do cultivar com inclinação máxima. Aplica-se este método aos dois conjuntos de regressões.

4.2 Método do Contorno Superior

4.2.1 Reparametrização e intersecção de rectas

Estabelecendo $\theta = \frac{\alpha}{\beta}$ passamos a ter

$$y = \alpha + \beta x = \beta(x + \theta)$$

onde $-\theta$ é a abcissa de intersecção da recta $\alpha + \beta x$ com o eixo dos XX.

Por outro lado

$$y' = cy = \beta'(x + \theta)$$

com $\beta' = c\beta$, vendo-se que θ é invariante. Dado o par de rectas $\beta_i(x + \theta_i)$, $i = 1, 2$ a respectiva intersecção é

$$x(1) = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2}$$

enquanto que a intersecção de $\beta_1(x + \theta_1)$ com $c\beta_2(x + \theta_2)$, é

$$x(c) = \frac{c\alpha_2 - \alpha_1}{\beta_1 - c\beta_2} = \frac{\alpha_2 - \frac{\alpha_1}{c}}{\frac{\beta_1}{c} - \beta_2}$$

vindo

$$\frac{dx(c)}{dc} = \frac{\alpha_2\beta_1 - \alpha_1\beta_2}{(\beta_1 - c\beta_2)^2} = \beta_1\beta_2 \frac{\theta_2 - \theta_1}{(\beta_1 - c\beta_2)^2}$$

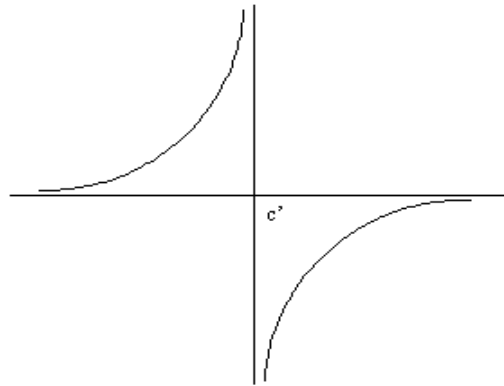
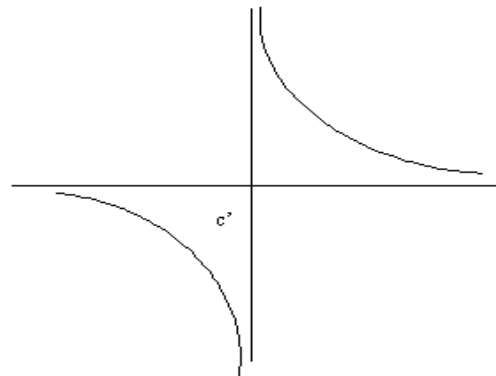
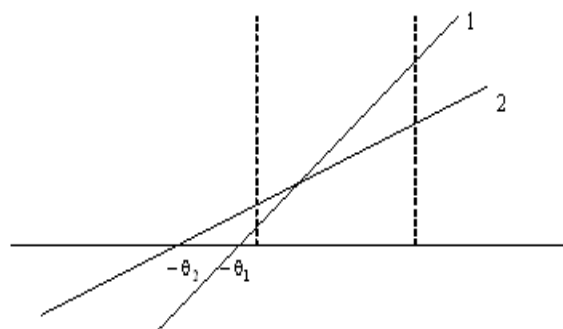
É interessante observar que ter $\frac{dx(c)}{dc} \geq 0$ é equivalente a $\theta_2 \geq \theta_1$ e por outro lado ter $\frac{dx(c)}{dc} \leq 0$ é equivalente a $\theta_2 \leq \theta_1$. Verifica-se que $\frac{dx(c)}{dc}$ tem sinal constante estando definida para $c \neq c' - \frac{\beta_1}{\beta_2}$, pois para este valor $x(c)$ e $x'(c)$ têm descontinuidade. Se $\theta_1 < \theta_2$, então $\frac{dx(c)}{dc} > 0$ e temos

Enquanto que, se $\theta_1 > \theta_2$

É imediato que $x(c) \rightarrow -\theta_2$ quando $c \rightarrow \infty$ e $x(c) \rightarrow -\theta_1$ quando $c \rightarrow 0$.

Ora neste caso interessa-nos sobretudo as intersecções em que $\beta_j > 0$, $j = 1, 2$, que se têm com $0 < \theta_{min} < x(c) < \theta_{max}$, sendo θ_{min} e θ_{max} respectivamente o mínimo e o máximo dos índices ambientais nas aplicações que vamos considerar. Se $\beta_1 > \beta_2$ e $\theta_1 \geq \theta_2$ ter-se-á $\alpha_2 < \alpha_1$ e $x(1) < 0$, pelo que a intersecção não se daria no intervalo $[u; v]$.

Assim para se ter $\theta_{min} \leq x(1) \leq \theta_{max}$ terá de se ter $\theta_1 < \theta_2$ vindo $\frac{dx(c)}{dc} > 0$, e

Figura 4.1: Gráfico da função $x(c)$ com $\theta_1 < \theta_2$ Figura 4.2: Gráfico da função $x(c)$ com $\theta_1 > \theta_2$ Figura 4.3: Representação gráfica da intersecção das rectas quando $\theta_1 < \theta_2$ 

Sejam c' e c'' tais que

$$\begin{cases} x(c') = \theta_{min} \\ x(c'') = \theta_{max} \end{cases}$$

então vê-se que se tem de ter

$$c' < c < c''$$

para que

$$\theta_{min} \leq x(c) \leq \theta_{max}.$$

Verifica-se ainda

$$\begin{cases} c' = \frac{\alpha_1 + \beta_1 \theta_{min}}{\alpha_2 + \beta_2 \theta_{min}} = \frac{\beta_1(\theta_{min} + \theta_1)}{\beta_2(\theta_{min} + \theta_2)} \\ c'' = \frac{\alpha_1 + \beta_1 \theta_{max}}{\alpha_2 + \beta_2 \theta_{max}} = \frac{\beta_1(\theta_{max} + \theta_1)}{\beta_2(\theta_{max} + \theta_2)} \end{cases}$$

logo para se ter

$$c' < 1 < c''$$

terá que se verificar

$$\frac{\beta_1}{\beta_2} \left(\frac{\theta_{min} + \theta_1}{\theta_{min} + \theta_2} \right) < 1 < \frac{\beta_1}{\beta_2} \left(\frac{\theta_{max} + \theta_1}{\theta_{max} + \theta_2} \right)$$

ou seja

$$\frac{\theta_{min} + \theta_1}{\theta_{min} + \theta_2} < \frac{\beta_2}{\beta_1} < \frac{\theta_{max} + \theta_1}{\theta_{min} + \theta_2}.$$

Podemos então afirmar que $c(\beta_2 x + \theta_2)$ e $\beta_1 x + \theta_1$ se intersectam em $[\theta_{min}; \theta_{max}]$ se e só se

$$\frac{\beta_1}{\beta_2} \left(\frac{\theta_{min} + \theta_1}{\theta_{min} + \theta_2} \right) < c < \frac{\beta_1}{\beta_2} \left(\frac{\theta_{max} + \theta_1}{\theta_{max} + \theta_2} \right).$$

Observe-se ainda que

$$\begin{aligned} c' - c'' &= \frac{\beta_1(\theta_{min} + \theta_1)}{\beta_2(\theta_{min} + \theta_2)} - \frac{\beta_1(\theta_{max} + \theta_1)}{\beta_2(\theta_{max} + \theta_2)} = \frac{\beta_1 \beta_2 (\theta_{min} + \theta_1)(\theta_{max} + \theta_2) - \beta_1 \beta_2 (\theta_{min} + \theta_2)(\theta_{max} + \theta_1)}{\beta_2^2 (\theta_{min} + \theta_2)(\theta_{max} + \theta_2)} \\ &= \frac{\beta_1}{\beta_2 (\theta_{min} + \theta_2)(\theta_{max} + \theta_2)} [\theta_{min} \theta_{max} + \theta_2 \theta_{min} + \theta_1 \theta_{max} + \theta_1 \theta_2 - \theta_{min} \theta_{max} - \theta_1 \theta_{min} - \\ &\quad \theta_2 \theta_{max} - \theta_1 \theta_2] \\ &= \frac{\beta_1}{\beta_2 (\theta_{min} + \theta_2)(\theta_{max} + \theta_2)} [\theta_2 - \theta_1 (\theta_{min} - \theta_{max})] \end{aligned}$$

Dado que a diferença $\theta_{min} - \theta_{max}$ é sempre negativa, ter $c' \geq c''$ é equivalente a $\theta_1 \geq \theta_2$ o que corresponde a ter-se $\frac{dx(c)}{dc} \leq 0$.

Por outro lado $c' \leq c''$ é equivalente a $\theta_1 \leq \theta_2$ o que significa ter-se $\frac{dx(c)}{dc} \geq 0$

4.2.2 Reparametrização e intersecção de várias rectas

Comecemos por admitir que $\beta_1 > \beta_2 > \dots > \beta_J$ e $\theta_1 < \min\{\theta_2, \dots, \theta_J\}$. As intersecções entre as rectas $\beta_1(x + \theta_1)$ e as $c_j\beta_j(x + \theta_j)$, $j = 2, \dots, J$ são dadas por

$$x_j(c) = \frac{c\alpha_j - \alpha_1}{\beta_1 - c\beta_j}, j = 2, \dots, J$$

verificando-se a condição $\frac{dx_j(c)}{dc} > 0$, $j = 2, \dots, J$.

Seja agora

$$\begin{cases} x_j(c'_j) = \theta_{min} \\ x_j(c''_j) = \theta_{max} \end{cases}, \quad j = 2, \dots, J$$

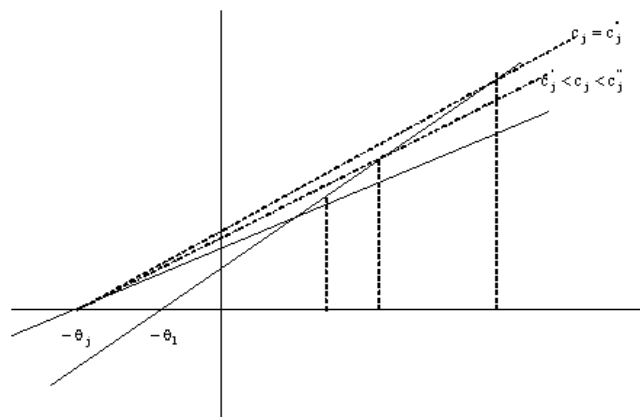
com $c'_j < c''_j$, $j = 2, \dots, J$.

Ora verifica-se que

$$\begin{cases} c'_j = \frac{\beta_1(\theta_{min} + \theta_1)}{\beta_j(\theta_{min} + \theta_j)} < \frac{\beta_1}{\beta_j} \\ c''_j = \frac{\beta_1(\theta_{max} + \theta_1)}{\beta_j(\theta_{max} + \theta_j)} < \frac{\beta_1}{\beta_j}, \end{cases} \quad j = 2, \dots, J$$

vindo $c'_j < c''_j < \frac{\beta_1}{\beta_j}$, $j = 2, \dots, J$.

Figura 4.4: Intersecção de várias rectas



Basta-nos então atender aos $a_j = \frac{c_j}{c'_j}$, $j = 2, \dots, J$ mantendo-se o cultivar 1 como dominante à direita desde que se tenha

$$\max\left\{\frac{c_j}{c'_j}, j = 2, \dots, J\right\} < 1$$

Observe-se que temos os ajustamentos iniciais c''_2, \dots, c''_J e a interacção adicional c_2, \dots, c_J , o que sugere a utilização da quase normalidade.

Admitindo que o cultivar que interessava foi "substituído", há que repetir o procedimento clássico com os coeficientes corrigidos para determinar os novos cultivares dominantes e os respectivos intervalos de dominância.

4.3 Comparações múltiplas para os coeficientes corrigidos

Passaremos a abordar nesta secção, o problema da comparação de cultivares numa perspectiva inovadora, na medida em que são introduzidos coeficientes de correcção.

Na secção 3.7.2 estabelecemos as comparações de cultivares em que sabendo-se os que são dominantes e conhecendo-se os respectivos coeficientes angulares ajustados, assim como o intervalo de dominância, procediamos a testes de comparação entre o primeiro cultivar dominante e os restantes e também entre o último cultivar dominante e os restantes.

Como referido no início do presente capítulo agora o principal objectivo é a selecção de cultivares com capacidade de resposta em índices ambientais altos. Assim bastará considerar os coeficientes estimados \tilde{V} e os produtos $\tilde{V}\tilde{\beta}\sim\beta\tilde{V}$.

Sejam $V_{1,j}, \dots, V_{m_j,j}$ os valores da variável complementar que se têm para o cultivar j , $j = 1, \dots, J$. Tomemos

$$V_{\bullet,j} = \frac{1}{m_j} \sum_{i=1}^{m_j} V_{i,j}, \quad j = 1, \dots, J$$

e

$$SQV = \sum_{j=1}^J \sum_{i=1}^{m_j} V_{i,j}^2 - \sum_{j=1}^J m_j V_{\bullet,j}^2$$

Admitamos agora que

- i) os valores da variável complementar têm variâncias $\hat{\sigma}^2$ e valores médios que dependem apenas do cultivar;

ii) $\hat{\sigma}^2$ é muito superior às $Var(\tilde{\beta}_j)$, $j=1, \dots, J$

Podemos então estimar $\hat{\sigma}^2$ através de

$$\tilde{\sigma}^2 = \frac{SQV}{\dot{g}}$$

com $\dot{g} = \sum_{j=1}^J m_j - J$,

e utilizar os resultados do capítulo 2 para tratar os produtos $\tilde{\beta}_j V_{\bullet,j}$, $j = 1, \dots, J$, como variáveis normais independentes com variâncias

$$\tilde{\beta}_j^2 \frac{\hat{\sigma}^2}{m_j}, j = 1, \dots, J$$

Dada a sua robustez utilizaremos o método de Scheffé para comparar estas variáveis. Então diferirão significativamente ao nível α os pares tais que

$$\frac{|\tilde{\beta}_j V_{\bullet,j} - \tilde{\beta}_{j'} V_{\bullet,j'}|}{\sqrt{(J-1) f_{1-\alpha, J-1, \dot{g}} \left(\frac{\tilde{\beta}_j^2}{m_j} + \frac{\tilde{\beta}_{j'}^2}{m_{j'}} \right) \tilde{\sigma}^2}} > 1 \quad (4.3.1)$$

Verifica-se que estes quocientes têm uma interessante propriedade de invariância.

Teorema 1 *Os quocientes $\frac{|\tilde{\beta}_j V_{\bullet,j} - \tilde{\beta}_{j'} V_{\bullet,j'}|}{\sqrt{(J-1) f_{1-\alpha, J-1, \dot{g}} \left(\frac{\tilde{\beta}_j^2}{m_j} + \frac{\tilde{\beta}_{j'}^2}{m_{j'}} \right) \tilde{\sigma}^2}}$ gozam da propriedade de invariância para transformações lineares.*

Dem: Começemos por considerar a substituição da variável V por $W = c + bV$. Nesse caso teremos

$$\hat{\sigma}^2 = Var(W) = b^2 Var(V) = b^2 \hat{\sigma}^2$$

assim como

$$W_{\bullet,j} = c + bV_{\bullet,j}, j = 1, \dots, J$$

vindo

$$\tilde{\sigma}^2 = b^2 \tilde{\sigma}^2$$

e teremos

$$\frac{|\tilde{\beta}_j W_{\bullet,j} - \tilde{\beta}_{j'} W_{\bullet,j'}|}{\sqrt{(J-1) f_{1-\alpha, J-1, \dot{g}} \left(\frac{\tilde{\beta}_j^2}{m_j} + \frac{\tilde{\beta}_{j'}^2}{m_{j'}} \right) \tilde{\sigma}^2}} = \frac{|\tilde{\beta}_j V_{\bullet,j} - \tilde{\beta}_{j'} V_{\bullet,j'}|}{\sqrt{(J-1) f_{1-\alpha, J-1, \dot{g}} \left(\frac{\tilde{\beta}_j^2}{m_j} + \frac{\tilde{\beta}_{j'}^2}{m_{j'}} \right) \tilde{\sigma}^2}} \text{ com } j \neq j'.$$

Assim é indiferente utilizar a variável complementar V ou a variável complementar W .

Capítulo 5

Análise de um Plano de Melhoramento de Aveia

5.1 Considerações Prévias

Na aplicação que se segue, pretendemos aplicar os resultados até aqui apresentados. Utilizaremos dados de produção de aveia, obtidos a partir de ensaios realizados pela Estação Nacional de Melhoramento de Plantas, nos anos de 1984 a 1999 onde se tinha 4 blocos por ensaio e um total de 11 cultivares por bloco e 2000,2002,2003 e 2004 com 2 blocos por ensaio e um total de 22 cultivares por bloco. Estamos assim a tratar um conjunto de dados bastante mais alargado daquilo que é usual em ACR. No plano estabelecido, optámos por uma análise ano a ano, permitindo-nos assim obter conclusões com mais pormenor contrariamente às obtidas numa análise global. Em cada ano há entrada e saída de cultivares, portanto a lista de cultivares não é fixa. Os locais dos ensaios também variam de ano de para ano. Como é usual nestes casos a utilização da ACR carece de validação. Iniciaremos o nosso estudo com a verificação de duas condições necessárias para essa validação.

- i) existência de correlação linear positiva elevada entre os índices ambientais ajustados;
- ii) existência de estabilidade dos cultivares ao longo do plano de melhoramento.

Para a verificação da primeira condição usaremos o coeficiente de correlação usual como medida da concordância entre os dois conjuntos de estimadores.

Tal como sugerido por Pinto(2006) se quiser testar a existência de concordância significativa, pode-se utilizar o coeficiente de correlação ordinal de Spearman.

Para ser viável a utilização da ACR dever-se-á ter alta concordância entre os dois conjuntos de estimadores para os índices ambientais assim como estabilidade para uma fracção considerável dos cultivares utilizados em mais dum ano. Devem portanto estas condições ser verificadas para o plano em análise. Proceder-se-á, à classificação e selecção dos cultivares, utilizando os resultados da análise ano a ano. Os cultivares serão divididos em dominados e dominantes e estes últimos, de acordo com a sua posição no seu intervalo de dominância, serão designados por dominantes à direita, intermédios ou dominantes à esquerda. Na análise ano a ano é possível identificar de forma muito clara, certas situações menos objectivas respeitantes à entrada e saída de cultivares do plano. Entre estas, é de considerar 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. 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 resultados. A introdução dos coeficientes de correcção permitirá obter resultados representativos não só da capacidade produtiva mas também do valor dessa produção. Apresentam-se de seguida a tabela 5.1 com lista dos cultivares ensaiados, a tabela 5.2 com indicação dos locais e anos onde foram efectuados os ensaios e a tabela 5.3 com indicação da presença[1]/ausência[0] dos cultivares nos vários anos.

Tabela 5.1: Lista dos cultivares ensaiados

AVON	S.MATEUS	SWAN	S.VICENTE
S.ROMÃO	ST.ALEIXO	AE8006	AE8007
AE8301	AE8302	AE8303	AE8401
AE8501	AE8602	AE8601	OBBER8132
AE8701	AE8801	AE8901	AE8902
AE9001	AE9002	AE9003	AE9004
AE9005	AE9101	AE9301	AE9302
AE9303	AE9401	AE9402	AE9403
AE9701	AE9702	AE9703	STA EULÁLIA
87SA29	COBER79-19	S.VICENTExAE8303	S.VICENTExCOBER
QR772	QR720	AVONxS.MATEUS	COBER79-19xAVON
X2795/LjuiL1743	TX88AB1494		

Tabela 5.2: Locais e anos onde se realizaram os ensaios

Experiência	Locais	Anos
1	Évora	1986,87,89,92,93,94,95,96,97,99,00,02,04
2	Coruche	1984,86,87,89
3	Herdade da Comenda	1988,94,96,97,99,00,03,04
4	E.N.M.P.	1984,85,86,87,88,89,90,91,92,93,94,95,97,98,99,00,02,03,04
5	Beja	1984,85,88,89,90,91,92,93,94,95,96,97,98,99,00,02,04
6	Benavila	1994,00,02
7	Portalegre	2003
8	Pegões	1990,91,92,93,95,96

Tabela 5.3: Presença/ausência dos cultivares nos vários anos

	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	02	03	04
AVON	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
S.MATEUS	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1
SWAN	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S.VICENTE	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
S.ROMÃO	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ST.ALEIXO	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AE8006	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8007	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8301	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8302	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8303	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8401	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8501	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
AE8602	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8601	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
OBER8132	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8701	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
AE8801	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
AE8901	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
AE8902	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
AE9001	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
AE9002	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0
AE9003	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	1	0	0	0
AE9004	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
AE9005	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
AE9101	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
AE9301	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
AE9302	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
AE9303	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
AE9401	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
AE9402	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
AE9403	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
AE9701	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
AE9702	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
AE9703	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
STA.EULÁLIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
87SA29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
COBER79-19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
S.VICENTExAE8303	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
S.VICENTExCOBER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
QR772	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
QR720	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
AVONxS.MATEUS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
COBER79-19xAVON	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
X2795/IJUiL1743	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
TX88AB1494	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1

Previamente convém referir que estamos a perante um plano de melhoramento onde se trabalha com um conjunto de cultivares que em cada ano se mantém fixo em todos os locais ensaiados. Os ensaios têm quatro repetições nos anos de 1984

a 1999 e duas repetições nos anos de 2000,2002,2003 e 2004. Neste caso temos uma rede que se designa de anualmente completa, para a qual podemos obter índices clássicos, calculando as médias dos blocos para cada ano. Apresentamos na tabela 5.4 os valores das médias dos índices ambientais L_2 anuais e índices clássicos obtidos

Tabela 5.4: Médias dos índices ambientais por bloco

Ano, Local	Ind. Amb. L2	Ind. Clássico	Ano, Local	Ind. Amb. L2	Ind. Clássico
1984, ENMP	786.217	795.461	1984, Beja	1744.976	1745.113
1984, Beja	1744.976	1745.113	1984, Coruche	465.904	469.909
1985, ENMP	1072.415	1054.642	1985, Beja	523.382	539.751
1986, Évora	651.587	659.955	1986, ENMP	1093.780	1098.430
1986, Coruche	215.744	217.364	1987, Évora	649.682	684.137
1987, ENMP	768.781	758.885	1987, Coruche	509.213	543.409
1988, Beja	569.264	613.501	1988, ENMP	495.576	484.546
1988, Elvas(H. de Comenda)	731.976	723.079	1989, Évora	847.771	846.489
1989, ENMP	893.761	891.212	1989, Coruche	91.443	91.572
1989, Beja	1484.040	1483.414	1990, ENMP	646.238	645.561
1990, Beja	548.247	547.213	1990, Pegões	79.796	80.560
1991, ENMP	1189.636	1192.584	1991, Beja	1262.827	1267.587
1991, Pegões	333.238	331.196	1992, ENMP	348.439	346.508
1992, Beja	447.291	442.370	1992, Évora	527.126	527.192
1992, Pegões	199.659	200.283	1993, ENMP	719.5679	723.130
1993, Évora	566.775	569.839	1993, Beja	1381.627	1375.475
1993, Pegões	153.131	152.665	1994, ENML	3617.923	3610.030
1994, Elvas(H. de Comenda)	3650.038	3636.364	1994, Benavila	4758.4803	4752.430
1994, Évora	838.163	843.611	1994, Beja	3107.144	3091.485
1995, ENML	1323.103	1305.061	1995, Évora	1403.569	1383.479
1995, Beja	1701.225	1687.860	1995, Pegões	539.789	541.396
1996, Elvas(H. de Comenda)	4718.569	4695.618	1996, Évora	3220.480	3227.814
1996, Beja	3102.389	3057.359	1996, Pegões	1809.380	1823.593
1997, ENMP	1381.841	1424.000	1997, Elvas (H. de Comenda)	2139.467	2131.841
1997, Évora	958.968	959.2274	1997, Beja	2231.379	2252.500
1998, ENMP	4068.375	4062.819	1998, Beja	1706.539	1707.455
1999, ENMP	2216.763	2251.614	1999, Elvas (H. de Comenda)	3239.127	3281.818
1999, Évora	4101.058	4138.227	1999, Beja	5067.539	5071.728
2000, ENMP	2435.816	2477.682	2000, Elvas(H. Comenda)	2589.281	1697.77
2000, Benavila	969.694	978.084	2000, Évora	1328.387	1323.051
2000, Beja	1462.88	1468.48	2002, ENMP	5106.903	5111.016
2002, Benavila	2456.913	2458.932	2002, Évora	2801.642	2848.89
2002, Beja	3162.086	3258.75	2003, ENMP	1818.332	1794.98
2003, Elvas(H. Comenda)	1769.112	1775.7	2003, Portalegre	776.482	792.41
2004, ENMP	4826.662	4816.75	2004, Elvas(H. Comenda)	2008.002	2009.045
2004, Évora	2307.344	2327.659	2004, Beja	2416.578	2436.5

Procedemos ao cálculo dos coeficientes de correlação linear usual e de Spearman tendo sido obtidos os seguintes resultados

Tabela 5.5: Soma dos quadrados dos Resíduos para a média dos índices ambientais

	Coeficiente de correlação de Spearman	Coeficiente de correlação linear
L_2 anual / Índices clássicos	0.9775	0.9855

Tabela 5.7: Coeficientes angulares ajustados (cont.)

Cultivares	1996	1997	1998	1999	2000	2002	2003	2004
AVON	0.933	0.919						
S.MATEUS			1.059	1.144	0.823	1.1442	1.218	0.98
SWAN								
S.VICENTE								
S.ROMÃO								
ST.ALEIXO	0.906	1.089	0.819	1.068	1.039	0.5982	1.146	1.114
AE8006								
AE8007								
AE8301								
AE8302								
AE8303								
AE8401								
AE8501								
AE8602								
AE8601								
OBER8132								
AE8701								
AE8801								
AE8901								
AE8902								
AE9001								
AE9002	1.255	0.884						
AE9003	0.915	1.457			0.937			
AE9004								
AE9005								
AE9101	0.947							
AE9301	0.728	0.682	1.026	0.886	1.085	1.277	0.731	1.033
AE9302	1.116	1.103	0.912	1.041	0.913	1.292	0.802	1.064
AE9303	0.970	1.010	0.943	0.881	0.974	1.042	1.685	0.926
AE9401	1.263	1.179	0.949	0.875	0.644	1.172	0.843	0.881
AE9402	0.991	0.891	1.202	1.236	1.106	0.994	0.626	0.724
AE9403	0.824	0.891	0.968	1.046	1.037	1.30	0.347	1.036
AE9701			1.104	0.927	0.888	0.837	0.347	0.709
AE9702			1.057	1.048	1.058	0.866	0.783	1.079
AE9703			0.934	0.772	0.973	1.004	0.771	1.186
STA.EULÁLIA					0.824	0.823	1.125	0.982
87SA29					1.148	0.694	1.952	1.39
COBER79-19					1.09	1.359	1.202	1.101
S.VICENTE \times AE8303					0.941	1.411	0.62	1.064
S.VICENTE \times COBER					1.119	1.01	0.793	0.923
QR772					0.995	0.566	1.514	1.072
QR720					1.337	0.602	1.491	0.985
AVON \times S.MATEUS					1.194	0.903	0.844	0.925
COBER79-19 \times AVON					0.939	0.834	0.53	0.82
X2795/IJUil1743					0.784	0.984	1.383	1.102
TX88AB1494						1.059	0.73	0.734

Segue-se a tabela contendo as somas dos quadrados dos resíduos obtidos em cada ano, onde se constata haver aqui uma certa uniformidade nos valores, pelo que será de considerar todos os cultivares presentes nas comparações a efectuar aos coeficientes angulares.

Tabela 5.8: Soma dos quadrados dos Resíduos para a média dos índices ambientais obtidos

Ano	1984	1985	1986	1987	1988	1989	1990
s	2.051×10^6	4.629×10^6	6.677×10^6	3.284×10^6	4.148×10^6	3.958×10^6	1.066×10^6
Ano	1991	1992	1993	1994	1995	1996	1997
s	5.379×10^6	2.187×10^6	3.190×10^6	7.865×10^7	1.142×10^7	4.855×10^7	3.214×10^7
Ano	1998	1999	2000	2002	2003	2004	
s	2.536×10^7	5.376×10^7	4.038×10^7	3.121×10^7	1.601×10^7	3.238×10^7	

Usaremos agora o estabelecido em 3.8 para realizar os testes correspondentes. Efectuados todos os cálculos necessários foram obtidos os seguintes resultados

Tabela 5.9: Resultados obtidos através dos testes F

Cultivar	Valor do teste F
AVON	0.5483
S.MATEUS	0.9871
SWAN	0.0035
S.VICENTE	0.2748
S.ROMÃO	0.7521
ST.ALEIXO	0.7949
AE8006	0.0371
AE8007	0.0462
AE8301	0.0421
AE8302	0.0039
AE8303	0.3994
AE8401	0.4764
AE8501	0.5884
AE8602	0.7165
AE8601	0.2768
OBER8132	(*)
AE8701	0.0018
AE8801	0.1774
AE8901	0.4128
AE8902	0.4714
AE9001	0.0305
AE9002	0.6218
AE9003	0.6179
AE9004	0.0211
AE9005	0.0181
AE9101	0.0001
AE9301	0.6833
AE9302	1.0840
AE9303	0.7610
AE9401	0.9233
AE9402	0.7344
AE9403	0.6863
AE9701	0.4280
AE9702	0.7967
AE9703	0.6246
STA.EULÁLIA	0.1338
87SA29	0.2491
COBER79-19	0.9191
S.VICENTE \times AE8303	0.8992
S.VICENTE \times COBER	0.0657
QR772	0.0761
QR720	0.0706
AVON \times S.MATEUS	0.2850
COBER79-19 \times AVON	0.0823
X2795/IJUil1743	0.0777
TX88AB1494	0.0241

Verifica-se haver apenas um caso, indicado por (*) num total de quarenta e seis cultivares em que existem diferenças significativas ao nível significância de 5%. Assim conclui-se estarmos em condições de proceder à análise através da ACR.

5.2 Análise do contorno superior

Aplicando o método do contorno superior atrás descrito obtivemos os resultados que se apresentam na tabela seguinte. Os gráficos dos contornos superiores que serviram de base para a análise efectuada encontram-se no Apêndice B. Neste caso assinalamos a presença dum cultivar com (P), cultivar dominante à esquerda com (E), cultivar dominante à direita com (D) e cultivar dominante intermédio com (I).

Tabela 5.10: Composição dos contornos superiores ano a ano

Cultivares	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	02	03	04
AVON	P	I	P	P	I	P	P	P	P	P	P	P	P	P						
S.MATEUS	P	D	P	P	P	P	D								P	P	P	P	P	P
SWAN	P	P																		
S.VICENTE	E	P	P	P					P	P										
S.ROMÃO	P	P	P	P	P	P	P													
ST.ALEIXO	P				P	P	P	P	P	P	E	P	P	E	P	P	P	P	P	P
AE8006	P	P	I																	
AE8007	I	P	E	P																
AE8301	P	P	P																	
AE8302	P	P	P																	
AE8303	D	E	D	P	E	P														
AE8401		P	P	P	P	P														
AE8501			P	P	P	E	P	P												
AE8602				D	D	P	P													
AE8601				P	I	D	E													
OBERS132				E	P															
AE8701					P	P	P													
AE8801						I	P	P	P	E										
AE8901							P	I	D	D										
AE8902							P	P	P	P										
AE9001							P	P	P	P										
AE9002							P	P	P	P	P	P	P							
AE9003							P	E	P	D	D	E	D				I			
AE9004							P	I	P	P										
AE9005							P	P	P	P										
AE9101											P	P	P	P						
AE9301											P	E	P	P	P	E	P	P	P	P
AE9302											P	P	P	P	P	P	P	P	P	P
AE9303											P	P	P	P	P	D	P	P	I	I
AE9401												P	D	P	P	P	P	P	P	P
AE9402												P	P	P	D	P	E	P	P	P
AE9403												P	P	P	P	I	P	P	E	P
AE9701																P	P	P	P	P
AE9702																P	P	P	P	P
AE9703																P	P	P	E	P
STA.EULÁLIA																	P	P	P	P
87SA29																		P	D	D
COBER79-19																		P	P	P
S.VICENTExAE8303																		P	P	P
S.VICENTExCOBER																		P	P	P
QR772																			I	P
QR720																			D	P
AVONxS.MATEUS																			P	P
COBER79-19xAVON																			P	D
X2795/IJUiL1743																			P	P
TX88AB1494																				P

Podemos desde já referir alguns aspectos relevantes da análise efectuada aos resultados.

Assim, verifica-se que os cultivares AVON(14 anos), S.Mateus(13 anos) e St.Aleixo(17 anos) foram os que permaneceram mais tempo na rede de ensaios, sendo que a média de permanência de todos os cultivares é aproximadamente 5.7 anos.

O cultivar S. Mateus dominante à direita em 1985 e em 1990 foi no entanto eliminado, tendo sido reintroduzido em 1998.

O cultivar AE8006 foi eliminado após ter sido dominante intermédio em 1986.

O cultivar AE8901 foi dominante à direita em 1992 e 1993 e foi eliminado.

O cultivar AE9003 foi dominante nos anos 1994, 1995, 1996 e 1997 e foi eliminado.

O cultivar AE9003 foi reintroduzido em 2000 foi dominante intermédio nesse ano e foi eliminado.

5.3 Contorno superior e selecção

A tabela 5.11 contém os coeficientes ajustados $\tilde{\alpha}$ e $\tilde{\beta}$, e os valores dos coeficientes de determinação R^2 para as regressões ajustadas referentes aos anos 1994, 1995, 1996 e 1997. A partir de agora iremos estudar com mais pormenor as regressões ajustadas respeitantes a estes anos, uma vez que aqui temos informação relativamente ao valor económico da produção. Será interessante fazer em primeiro lugar uma análise clássica do problema e posteriormente fazer a introdução do método através da utilização dos coeficientes angulares corrigidos.

Tabela 5.11: Resultados obtidos no ajustamento

	1994			1995			1996			1997		
Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	R^2	$\tilde{\alpha}$	$\tilde{\beta}$	R^2	$\tilde{\alpha}$	$\tilde{\beta}$	R^2	$\tilde{\alpha}$	$\tilde{\beta}$	R^2
AE9003	13.939	1.248	0.920	-79.201	1.235	0.666	886.51	0.914	0.883	-186.81	1.457	0.89
AE9101	-85.038	0.969	0.899	-249.908	1.210	0.837	63.97	0.947	0.848	-42.37	0.943	0.917
AE9401				-72.591	1.052	0.904	-1079.91	1.263	0.943	-356.53	1.179	0.931
AE9302	170.78	0.948	0.933	-24.321	1.010	0.778	-197.36	1.117	0.844	-14.22	1.103	0.865
ST ALEIXO	196.52	0.886	0.757	32.114	1.004	0.823	679.48	0.906	0.874	220.15	1.089	0.664
AE9303	-122.87	1.082	0.869	0.116	1.010	0.771	67.49	0.97	0.908	-111.66	1.01	0.81
AE9002	-231.79	0.757	0.963	67.513	0.972	0.813	-595.26	1.255	0.891	374.19	0.884	0.757
AE9301	168.33	1.000	0.910	130.786	0.910	0.805	835.84	0.728	0.75	234.15	0.682	0.733
AVON	90.600	0.928	0.876	73.456	0.903	0.711	128.66	0.933	0.821	-69.67	0.919	0.582
AE9402				68.185	0.872	0.725	-156.39	0.991	0.89	-32.65	0.891	0.88
AE9403				73.456	0.713	0.660	-269.83	0.824	0.89	-40.4	0.891	0.891
AE9001	9.94	0.911	0.823									
AE9004	207.1	0.899	0.848									
AE9005	-342.74	1.005	0.924									

Tendo em atenção os valores apresentados para R^2 , podemos considerar que estas regressões estão bem ajustadas.

As rectas de regressão estão representadas graficamente, por ano, no Apêndice B,

sendo possível observar quais são os cultivares dominantes à esquerda, à direita e intermédios.

É importante agora proceder a testes de comparação múltipla para verificar da existência de cultivares significativamente dominados. Conhecidos os cultivares dominantes e os intervalos de dominância nos anos considerados vamos em primeiro lugar proceder à realização de testes t unilaterais ao nível de significância de 5%, de acordo com o estabelecido em 3.7.2. obtiveram-se os seguintes resultados

Tabela 5.12: Cultivares dominantes e dominados: resultados da aplicação do teste t

Cultivar dominante	Intervalo de dominância	Cultivares dominados ao nível de 5%
Ano 1994		
AE9003	[1717.76 ; 5032.46]	AE9101,AE9303,AE9301,AE9302 AVON,ST.ALEIXO,AE9001,AE9005 AE9004,AE9002
St Aleixo	[642.86 ; 1717.76]	AE9003,AE9101,AE9303 AE9301,AE9302,AVON
Ano 1995		
AE9003	[647.34 ; 1881.31]	AE9401,AE9302,ST.ALEIXO,AE9303 AE9002,AE9301,AVON,AE9402,AE9403
AE9301	[380.41 ; 647.34]	AE9401,AE9302,ST.ALEIXO AE9303,AE9002
Ano 1996		
AE9002	[4347.98 ; 4897.19]	AE9402,AE9303,AE9101 AVON,ST.ALEIXO
AE9003	[1634.2 ; 4347.98]	AE9002,AE9302
Ano 1997		
AE9003	[1107.21 ; 2836.36]	AE9302,ST.ALEIXO,AVON,AE9002
St Aleixo	[813.82 ; 1107.21]	AE9003,AE9401,AE9302

Apresentam-se agora na tabela 5.13 os resultados obtidos através dos métodos de comparação múltipla de Scheffé e de Bonferroni. O método de Scheffé sendo um método bastante conservador, tal como seria de esperar apresenta uma grande percentagem de casos onde predomina a ausência de diferenças significativas. Apenas no ano de 1996 se verifica que o cultivar AE9003 domina no respectivo intervalo o cultivar AE9403, enquanto o mesmo se passa relativamente ao cultivar AE9002 sobre o cultivar AE9302, AE9101 e AE9401.

Tabela 5.13: Resultados dos métodos de comparação múltipla de Scheffé e de Bonferroni ao nível de 5% - Anos 1994,1995,1996,1997

	1994	1994	1995	1995	1996	1996	1997	1997
Método	St Aleixo	AE9003	AE9301	AE9003	AE9003	AE9002	St Aleixo	AE9003
Scheffé	nenhum	nenhum	nenhum	nenhum	AE9403	AE9302 AE9101 AE9401	nenhum	nenhum
Bonferroni	AE9303	AE9303 AE9001	AE9101	AE9101 AE9403	AE9101 AE9401 AE9302	AE9101 AE9401 AE9302 AE9402 AE9403	AE9401	AE9101 AE9401 ST.ALEIXO AE9303 AE9002 AE9301 AVON AE9403

5.4 Classificação e selecção dos cultivares com coeficientes angulares corrigidos

Finalmente vamos aplicar o método aproximado de comparações múltiplas de Scheffé estabelecido na secção 4.3., tendo presente que se verifica a propriedade de invariância do método para transformações lineares das variáveis complementares.

Agora serão considerados os coeficientes angulares e os coeficientes de correcção associados às produções. Tal como referido, agora interessa a selecção de cultivares com capacidade de resposta em índices ambientais altos. Segue-se a tabela 5.14 onde se mostram para os anos estudados os cultivares em análise, assim como os valores de m_j , $V_{\bullet,j}$, $\tilde{\beta}_j V_{\bullet,j}$, $j = 1, \dots, J$ e SQV , obtidos de acordo com o estabelecido na secção 4.3 e onde Q traduz os valores dos quocientes das desigualdades 4.3.1 .

Tabela 5.14: Comparações múltiplas com coeficientes corrigidos

Cultivar	1994				1995				1997			
	m_j	$V_{\bullet,j}$	$\hat{\beta}_j V_{\bullet,j}$	Q	m_j	$V_{\bullet,j}$	$\hat{\beta}_j V_{\bullet,j}$	Q	m_j	$V_{\bullet,j}$	$\hat{\beta}_j V_{\bullet,j}$	Q
AE9003	20	42.58	42.79		16	32.31	39.90		4	41.18	60.00	
AE9101	20	43.75	42.39	0.09	16	35.68	43.17	0.38	4	41.80	39.42	3.26
AE9401					16	33.28	35.01	0.61	4	43.47	51.25	1.26
AE9302	20	41.74	39.57	0.69	16	31.91	32.23	0.97	4	42.50	46.88	1.95
ST ALEIXO	20	43.11	38.11	1.03	16	34.13	34.27	0.72	4	43.39	47.25	1.90
AE9303	20	44.78	48.45	1.13	16	34.43	34.77	0.65	4	42.85	43.28	2.58
AE9002	20	44.5	49.4	1.30	16	33.96	33.01	0.89	4	43.49	38.45	3.49
AE9301	20	43.63	40.84	0.42	16	33.27	30.28	1.27	4	43.01	29.33	5.36
AVON	20	42.52	39.46	0.72	16	33.59	30.33	1.27	4	42.34	38.91	3.37
AE9402					16	32.44	28.29	1.56	4	41.06	36.58	3.78
AE9403					16	32.64	23.30	2.36	4	43.34	38.62	3.45
AE9001	20	41.16	37.5	1.15								
AE9004	20	43.25	40.14	0.57								
AE9005	20	42.52	53.06	1.89								
SQV		2547.7				3404.7				105.52		

Após aplicação do método corrigido apresentam-se na tabela 5.15 os cultivares dominantes à direita e os significativamente dominados por estes.

Tabela 5.15: Resultados do método aproximado de comparação múltipla de Scheffé ao nível de 5% - Anos 1994,1995,1996,1997

	1994	1995	1997
cultivar dominante	AE9003	AE9003	AE9003
cultivares dominados	ST ALEIXO	AVON	AE9101
	AE9002	AE9403	AE9401
	AE9001	AE9301	AE9302
	AE9303	AE9402	ST ALEIXO
	AE9005		AE9303
			AE9002
			AE9301
			AVON
		AE9402	
		AE9403	

Note-se que nestes três anos o cultivar agora dominante à direita é o AE9003. Embora nem todos os outros cultivares estejam significativamente dominados nos anos de 1994 e 1995, para índices ambientais altos, a permanência do AE9003 na posição cimeira é um resultado de grande interesse prático. Este cultivar é

eliminado em 1997 o que é difícil de justificar.

5.5 Notas críticas

Vimos na alínea anterior ser difícil de justificar a eliminação do cultivar AE9003 a partir de 1997, tanto mais que em 1997 conclui-se que domina quase todos os outros cultivares, quando aplicamos o método de Bonferroni e domina todos os outros cultivares quando aplicamos o método de Scheffé aproximado. Além disso é também difícil de aceitar:

- i)* a eliminação do cultivar AE8901 em 1993 quando tinha sido dominante à direita.
- ii)* a não eliminação do cultivar AE9403 que é significativamente dominado em três anos consecutivos.

Positivamente vemos as eliminações efectuadas aos cultivares AVON em 1997 e AE9002 também em 1997. Face a isto vê-se que a ACR e a ACR completada que podem ser aplicadas ano a ano, constituem instrumentos válidos para a condução de planos de melhoramento.

Esta também era a conclusão para a ACR de Pinto(2006) na análise do plano de melhoramento de trigo.

Observe-se ainda que nos anos estudados a comparação dos βV revelou maior poder separador que os métodos de Scheffé e de Bonferroni.

Capítulo 6

Desenvolvimentos futuros

Embora a técnica desenvolvida tenha interesse prático e fosse perfeitamente válido aplicá-la a outras culturas é verdade que a mesma foi possível devido a ter-se podido mostrar que o produto de duas variáveis normais é, desde que uma delas tenha coeficiente de variação pequeno, aproximadamente normal.

Aliás a normalidade aproximada tem recentemente permitido tratar problemas que doutra forma seriam pouco manejáveis, ver por exemplo Ramos(2007).

Assim, tenciono dedicar-me ao desenvolvimento da quase-normalidade continuando a colaborar com o meu orientador.

Apêndice A

Introdução aos α -Designs

Inicialmente os delineamentos usados nas experimentações através de redes de ensaios eram do tipo de blocos casualizados. Os delineamentos em blocos incompletos tiveram a sua grande divulgação e utilização a partir da década de 70, concretamente através da vasta aplicação nos ensaios de variedades realizados no Reino Unido. Actualmente na Europa, os α -designs são os delineamentos mais usados tendendo a ser cada vez mais utilizados na comparação de cultivares.

Utilizando-se blocos incompletos é conveniente considerarmos delineamentos nos quais os blocos se agrupam em super-blocos. Nestes delineamentos os blocos contêm apenas uma parte dos cultivares, enquanto que os super-blocos deverão conter uma vez cada cultivar. Nestas condições, com $\alpha = 1$, podemos afirmar que estes delineamentos são resolúveis no sentido de Shrikhande & Raghavarao(1963 e 1964).

Em Patterson & Williams (1976) é desenvolvido o novo tipo de delineamentos em blocos incompletos - os α -designs. Sendo usado um processo cíclico na sua construção, estes delineamentos apresentam a grande vantagem de serem flexíveis. Em geral os α -designs podem ser usados sempre que o número de cultivares J seja múltiplo da dimensão nb dos super-blocos e nesse caso, cada cultivar estará presente α vezes em cada super-bloco.

É no entanto usual considerar-se $\alpha = 1$ por duas razões fundamentais:

- *permite a redução da dimensão dos super-blocos*
- *aumenta a homogeneidade no interior dos super-blocos*

A construção destes delineamentos foi no início dificultada pela falta de recursos computacionais que progressivamente foram surgindo. Patterson, Williams & Hunter (1978) apresentam um catálogo dos α -designs mais eficientes, chegando

a 100 parcelas por bloco e 4 super-blocos. Mais tarde são criados potentes algoritmos capazes de gerar quaisquer planos com factores de eficiência bastante elevados, por exemplo em Paterson & Patterson (1984) e em Patterson, Wild & Williams (1988).

Os avanços tecnológicos proporcionaram então uma notória evolução, nomeadamente no campo dos algoritmos, tornando possível a construção dos mais variados tipos de delineamentos. Actualmente a tarefa principal passa a ser o da escolha das melhores soluções tendo em atenção o problema que se pretende resolver.

Baseado em John & Williams (1995) foi desenvolvido no CSIRO (Commonwealth Scientific & Industrial Research Organisation) por D. Whitaker, E.R. Williams e J.A. John, um dos programas computacionais que permitem a construção de muitas classes de delineamentos. Este software permite a geração de α -designs utilizando um processo aleatório e atribui a cada delineamento gerado o respectivo factor de eficiência. Este processo facilita a procura de um delineamento óptimo a adaptar a diferentes realidades práticas. Informações detalhadas acerca do CycDesigN encontram-se disponíveis em www.ffp.csiro.au/tigr/software/cycdesign/ ou www.cycdesign.co.nz/. Mais recentemente, D. Whitaker, E.R. Williams e J.A. John apresentam o trabalho "CycDesigN Version 2.0: a package for the computer generation of experimental designs", correspondente a uma nova versão que permite a construção de delineamentos para o caso espacial.

Apêndice B

Gráficos com rectas ajustadas

Figura B.1: Análise conjunta de regressões - 1984

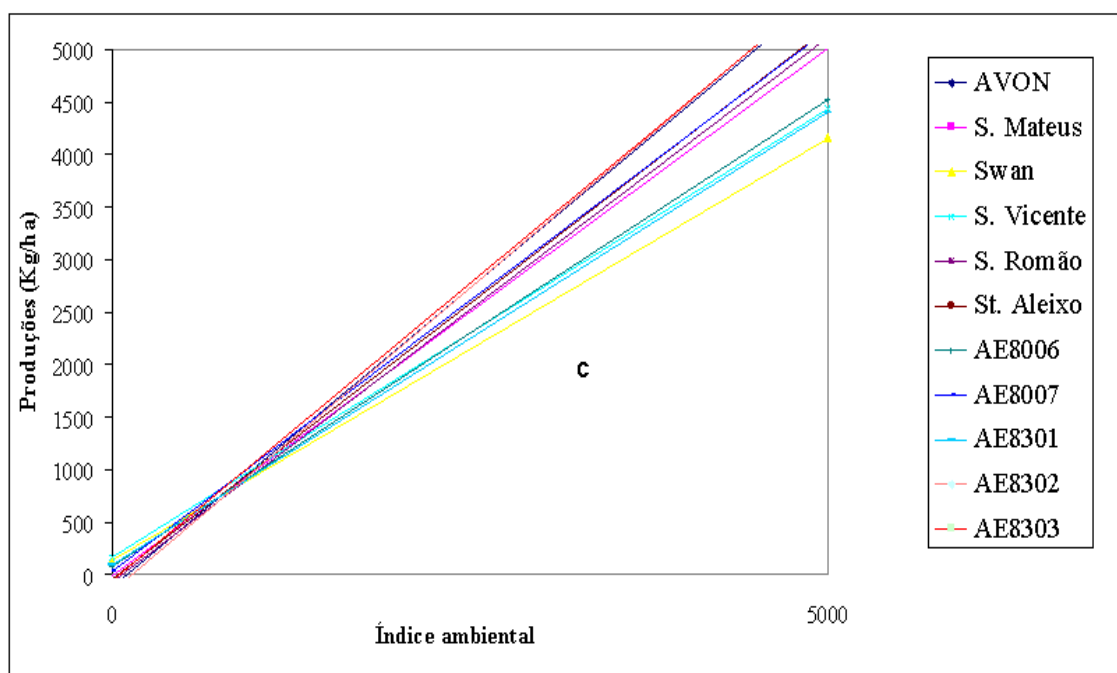


Figura B.2: Análise conjunta de regressões - 1985

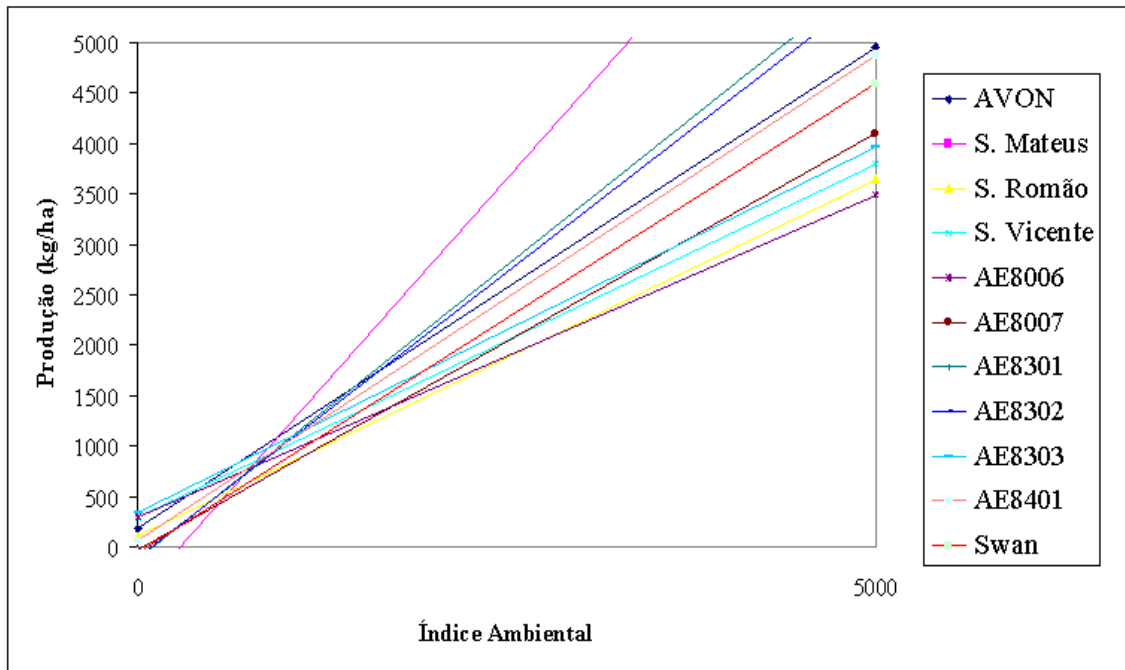


Figura B.3: Análise conjunta de regressões - 1986

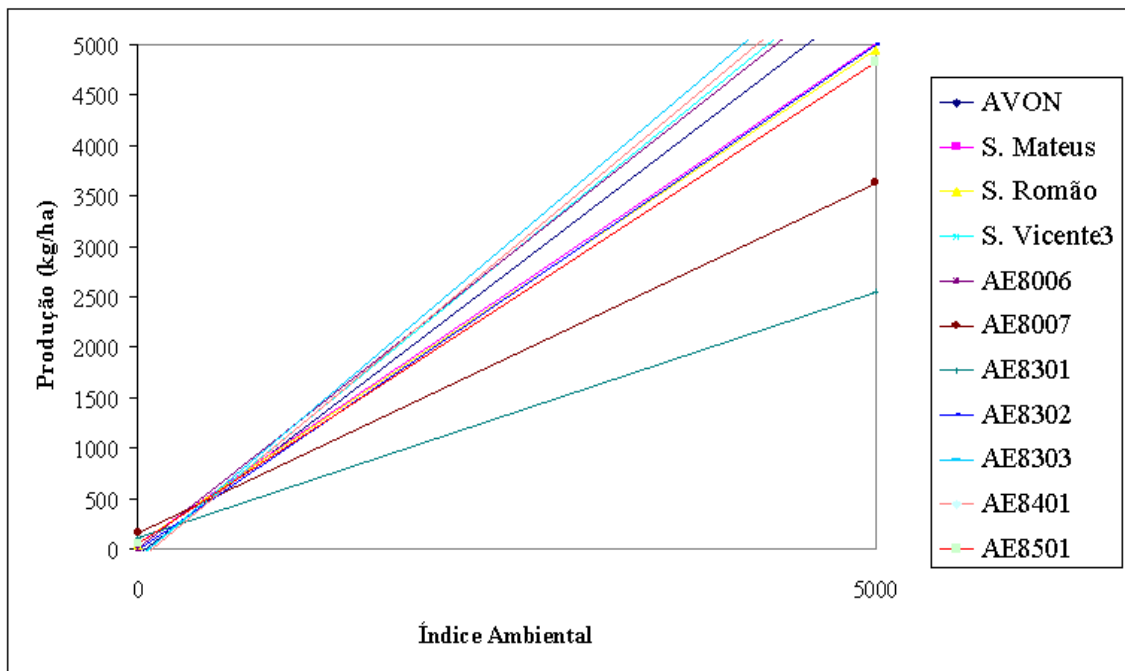


Figura B.4: Análise conjunta de regressões - 1987

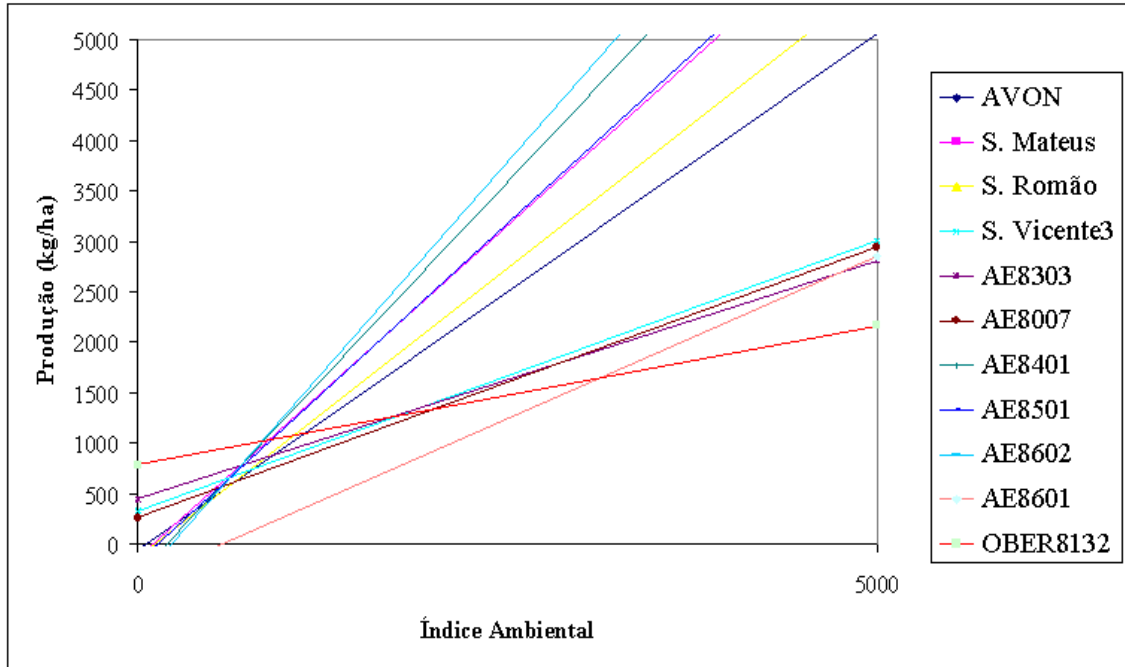


Figura B.5: Análise conjunta de regressões - 1988

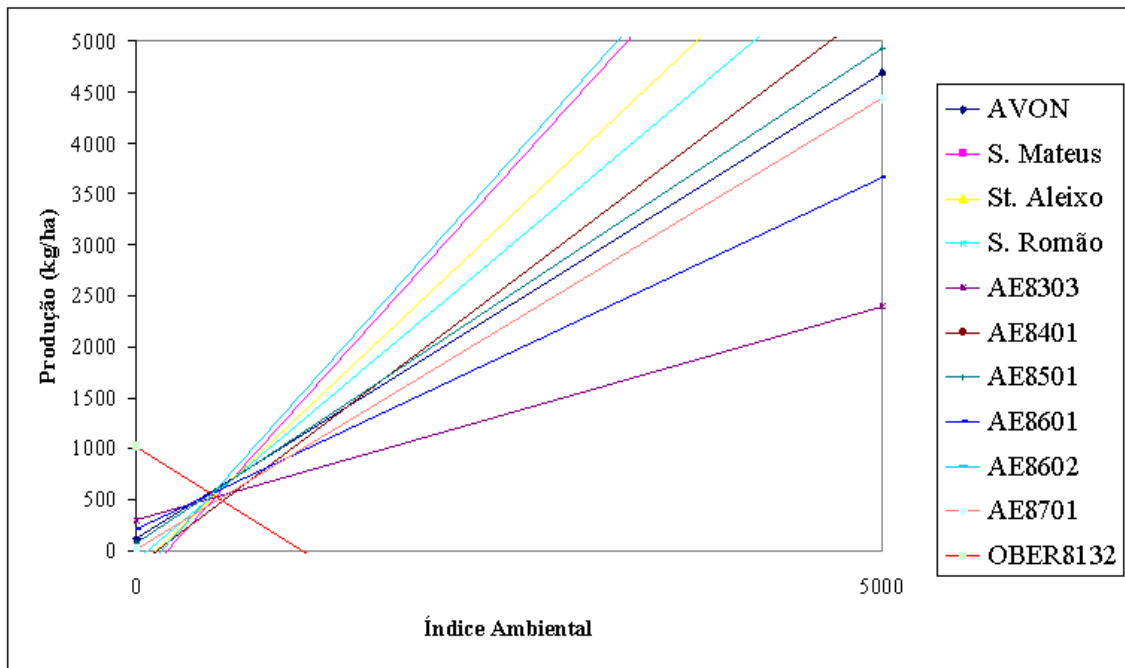


Figura B.6: Análise conjunta de regressões - 1989

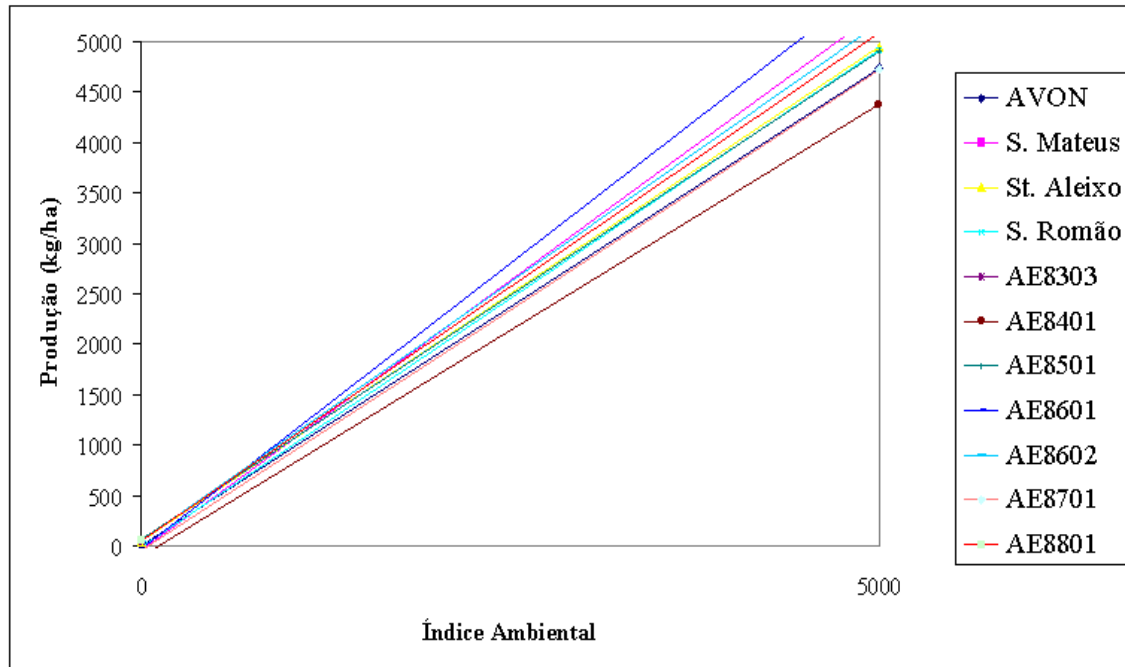


Figura B.7: Análise conjunta de regressões - 1990

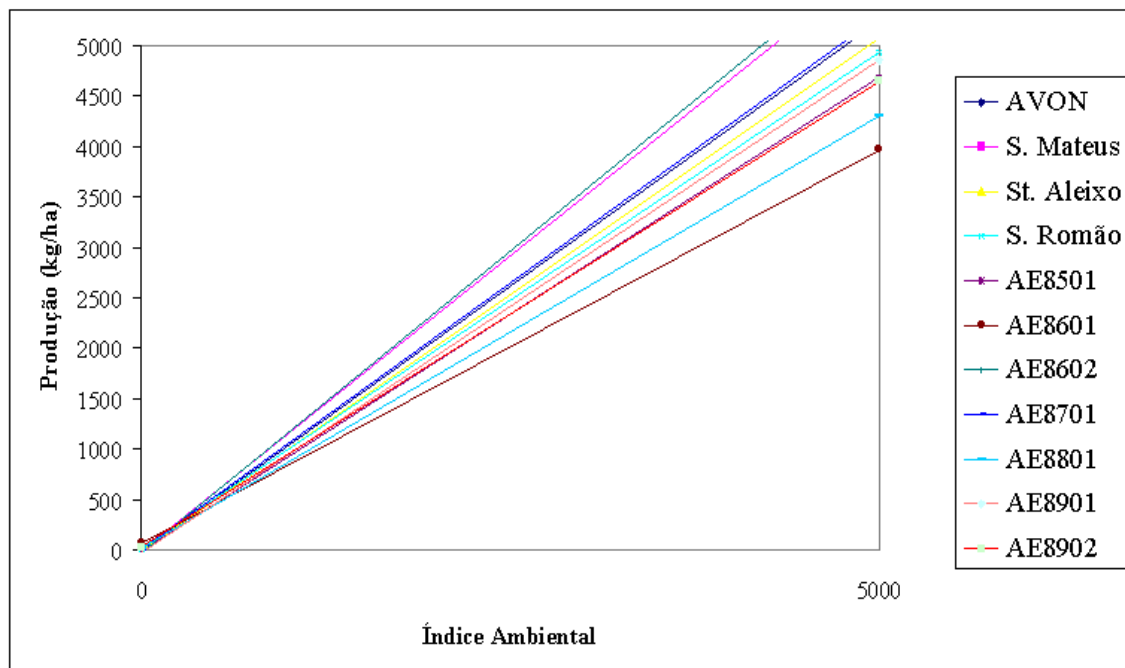


Figura B.8: Análise conjunta de regressões - 1991

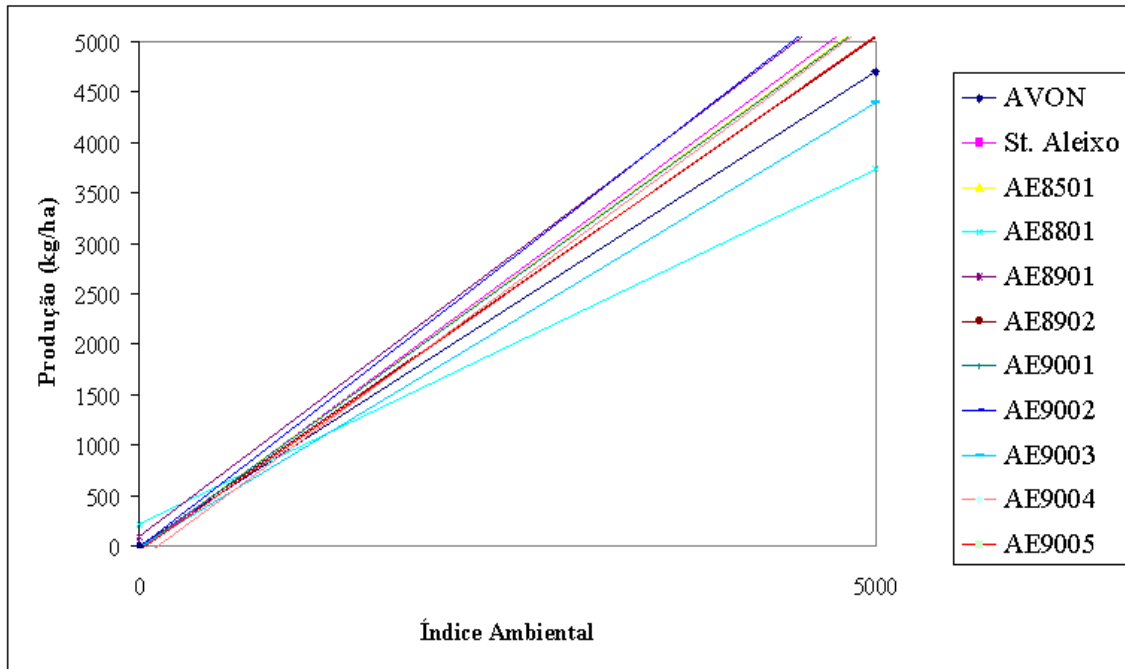


Figura B.9: Análise conjunta de regressões - 1992

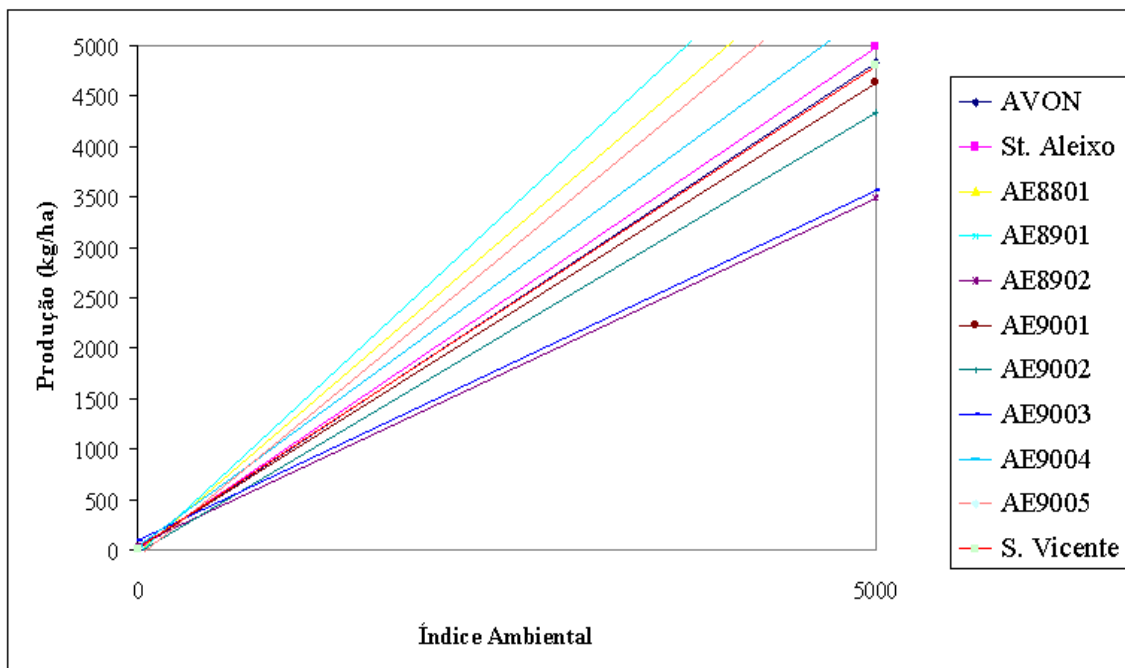


Figura B.10: Análise conjunta de regressões - 1993

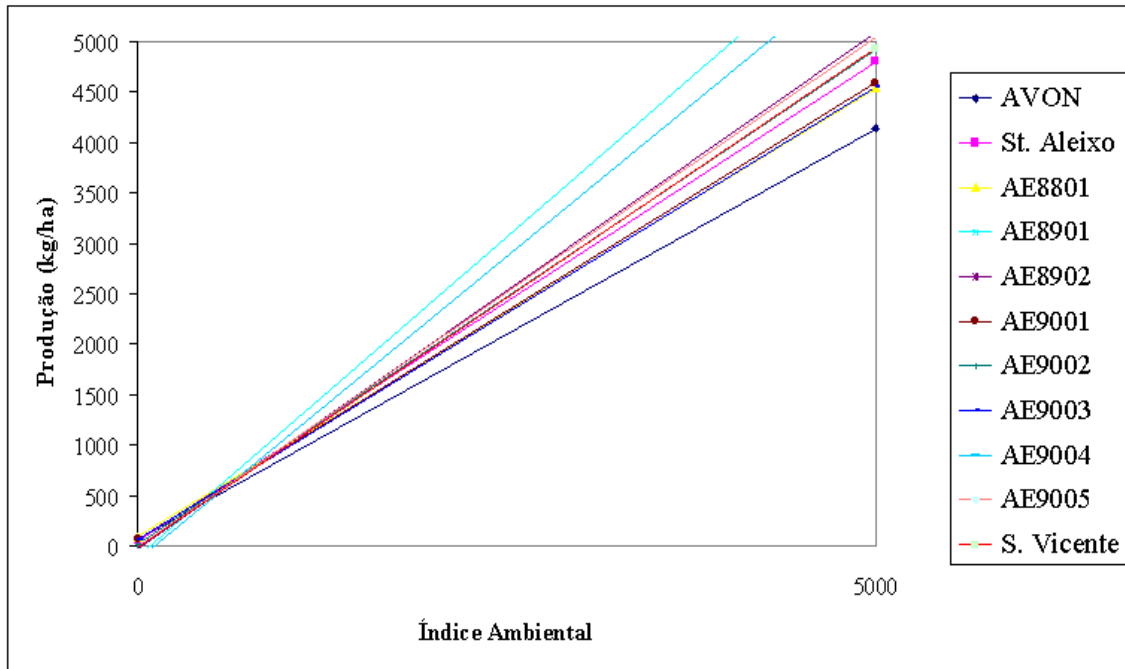


Figura B.11: Análise conjunta de regressões - 1994

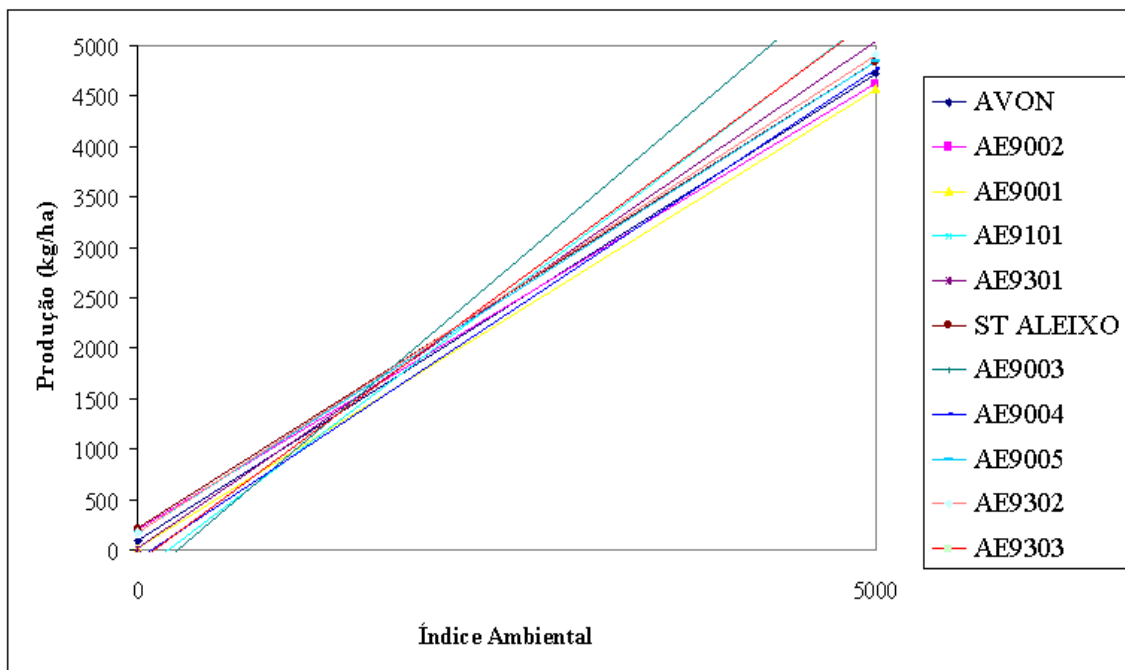


Figura B.12: Análise conjunta de regressões - 1995

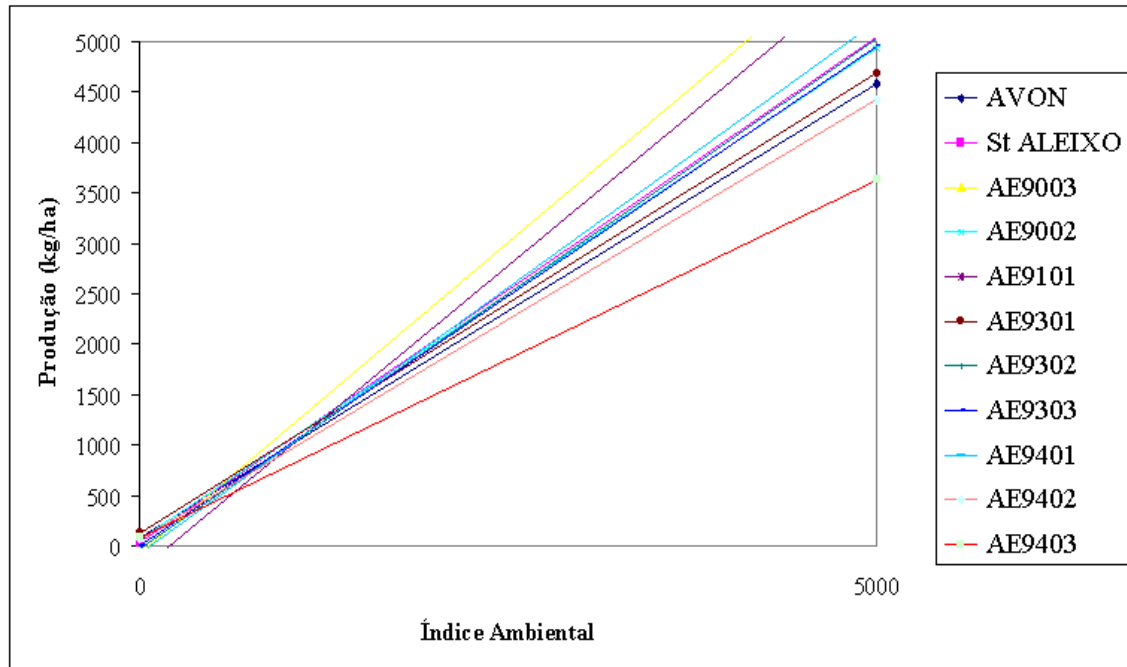


Figura B.13: Análise conjunta de regressões - 1996

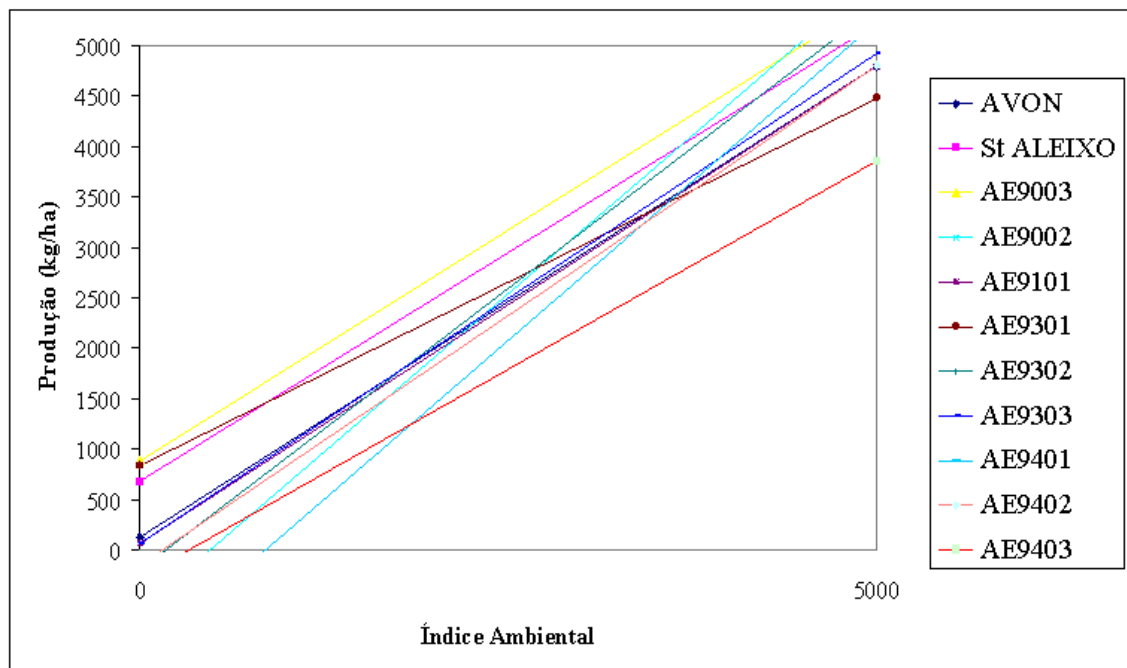


Figura B.14: Análise conjunta de regressões - 1997

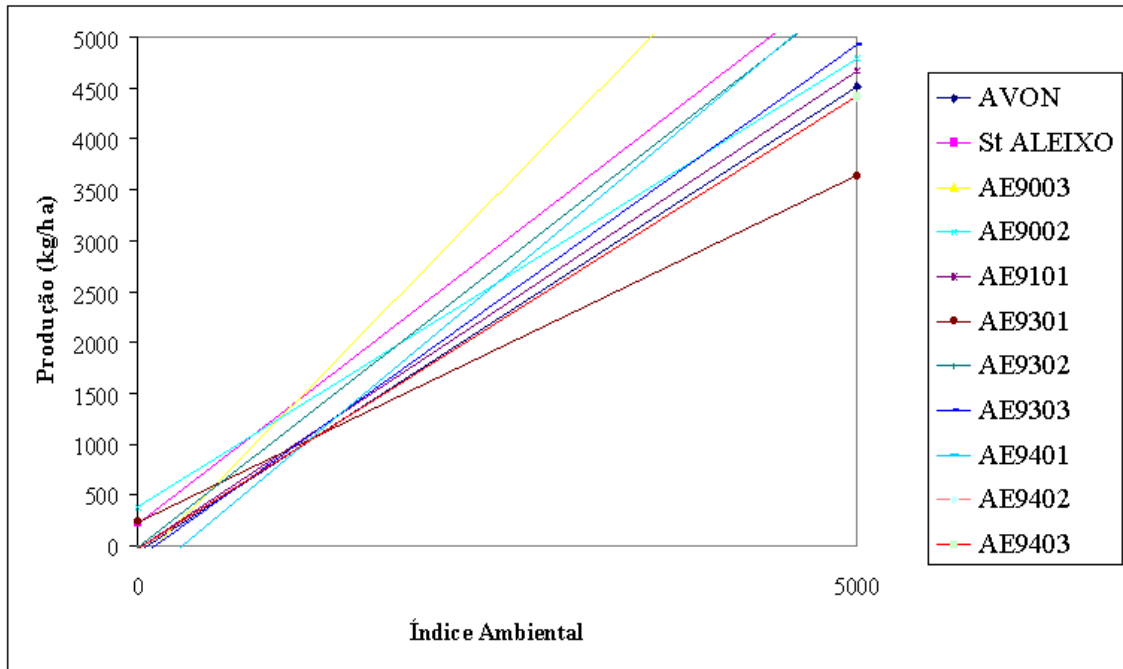


Figura B.15: Análise conjunta de regressões - 1998

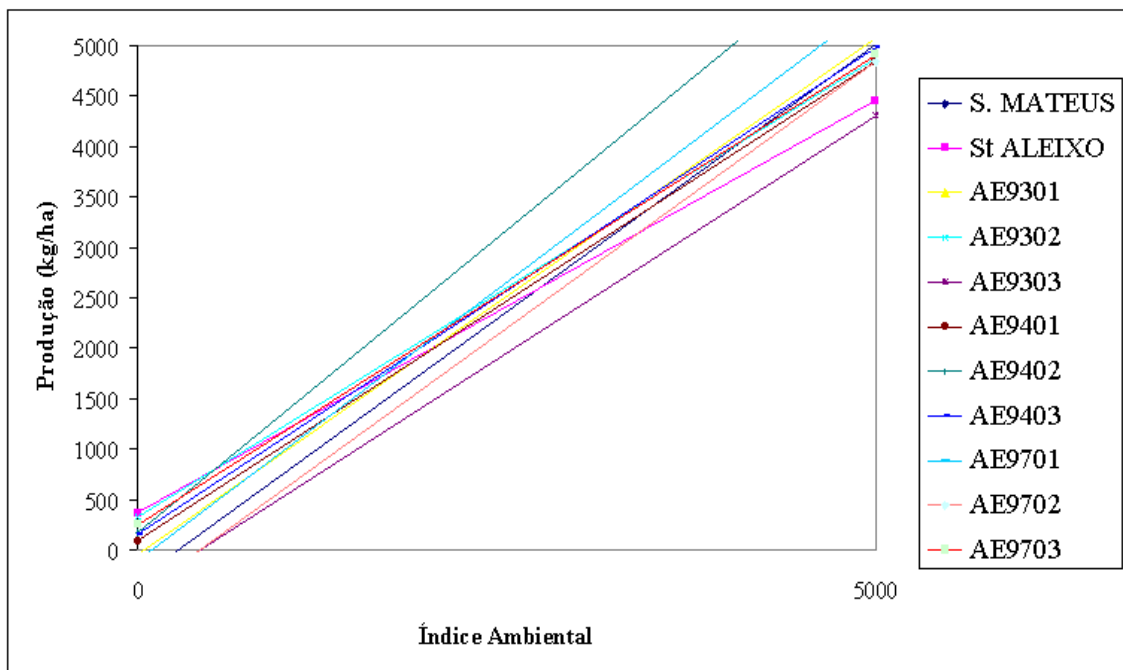


Figura B.16: Análise conjunta de regressões - 1999

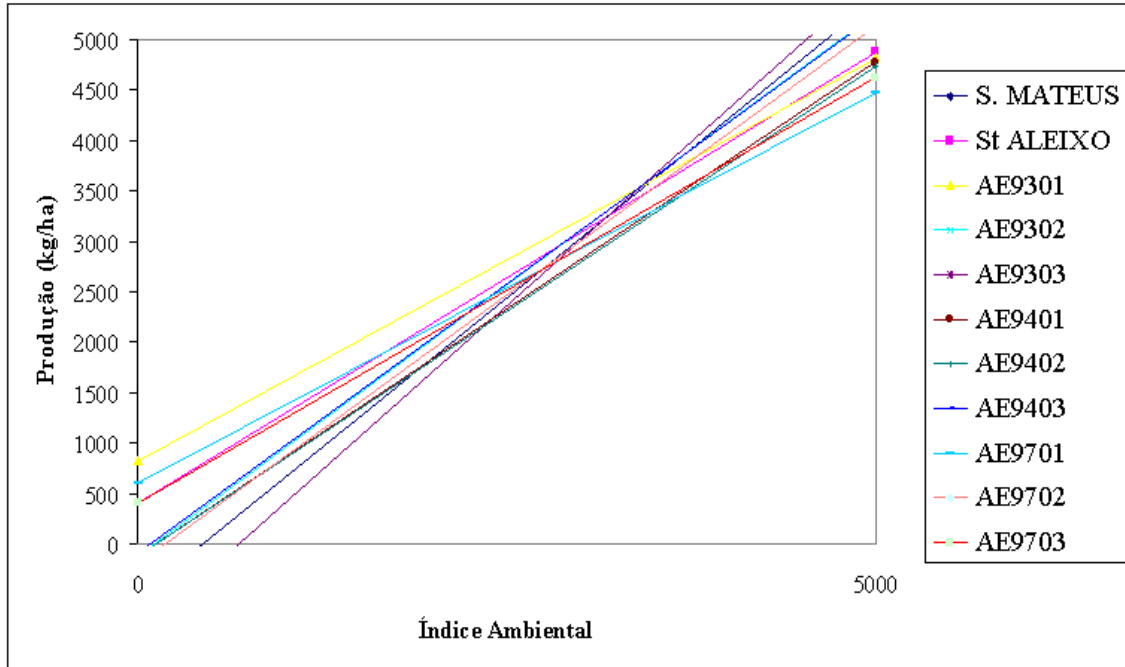


Figura B.17: Análise conjunta de regressões - 2000

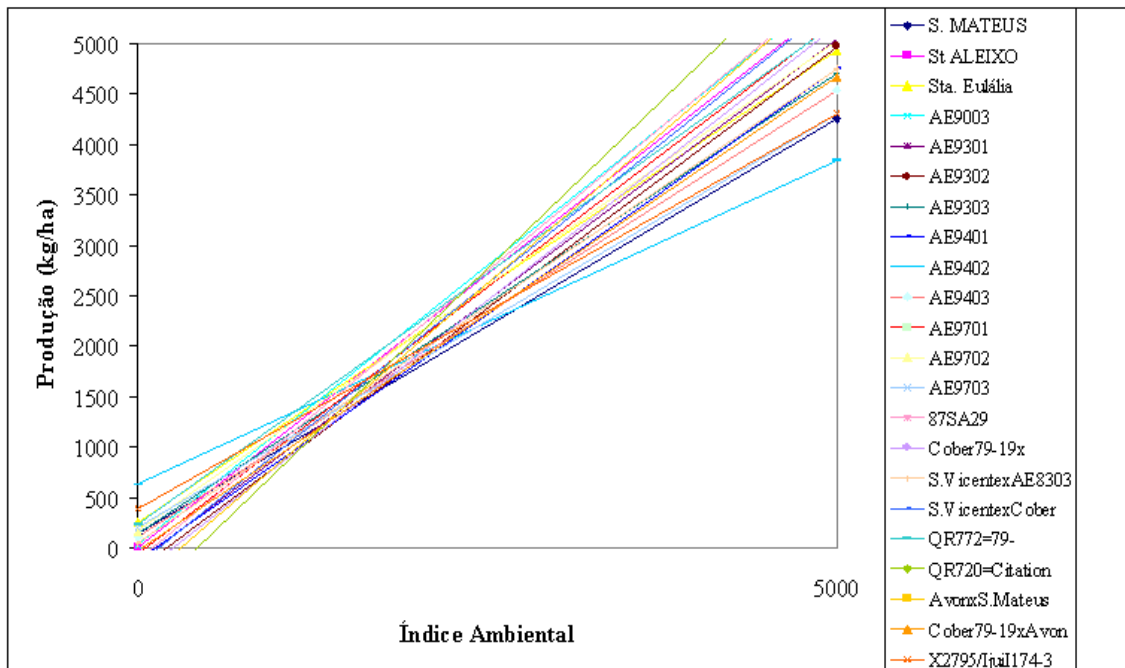


Figura B.18: Análise conjunta de regressões - 2002

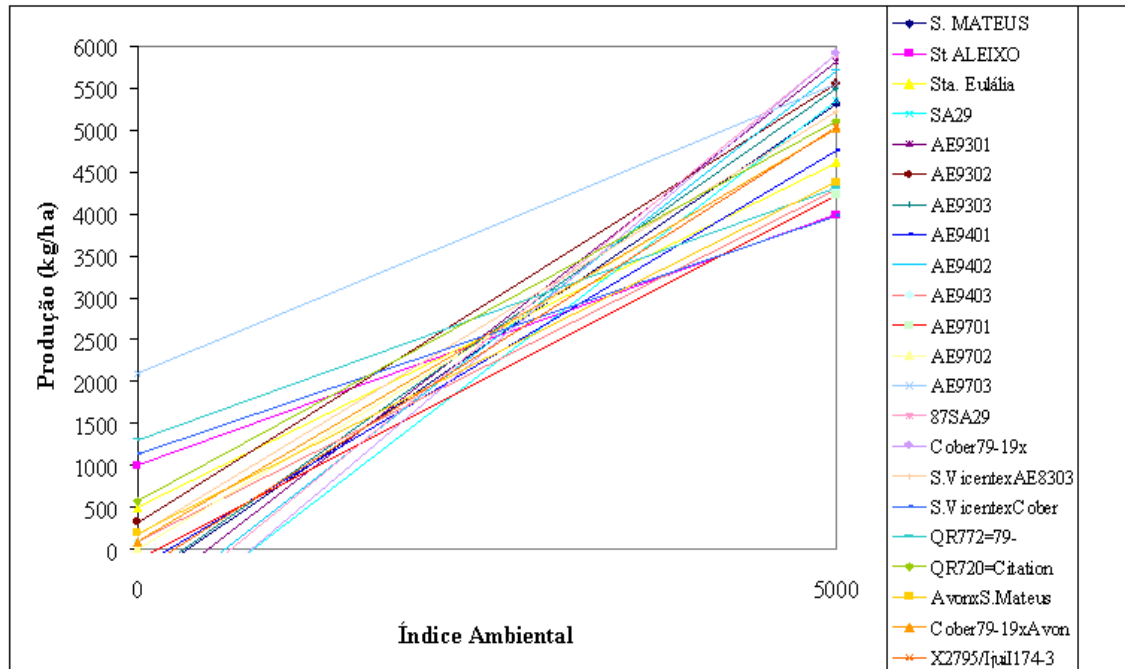


Figura B.19: Análise conjunta de regressões - 2003

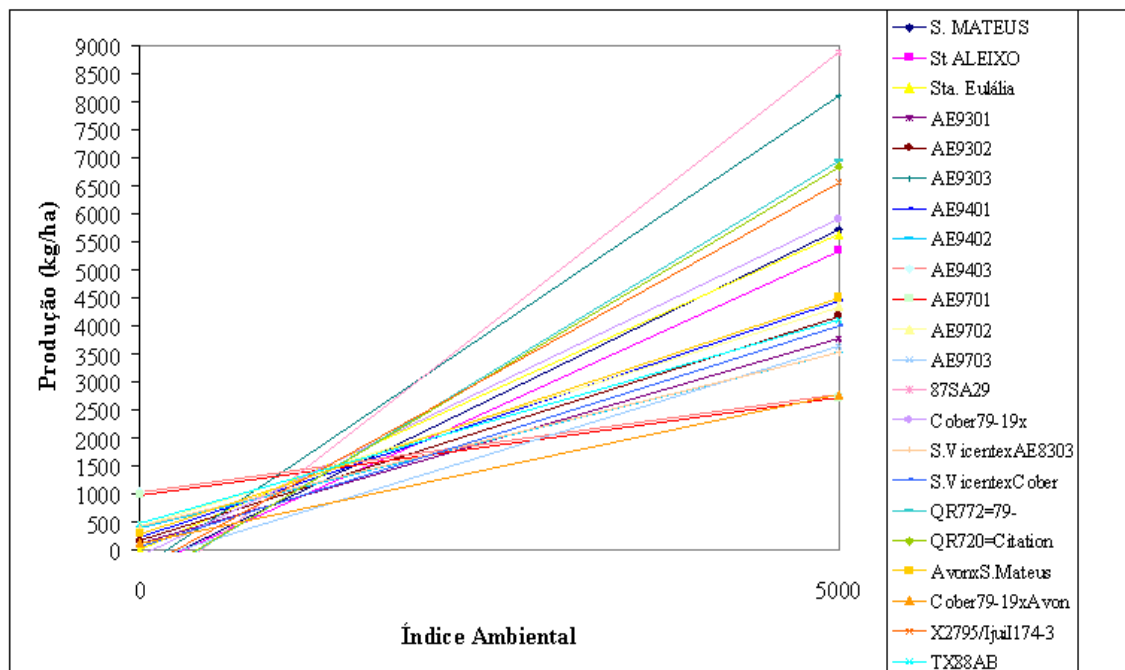
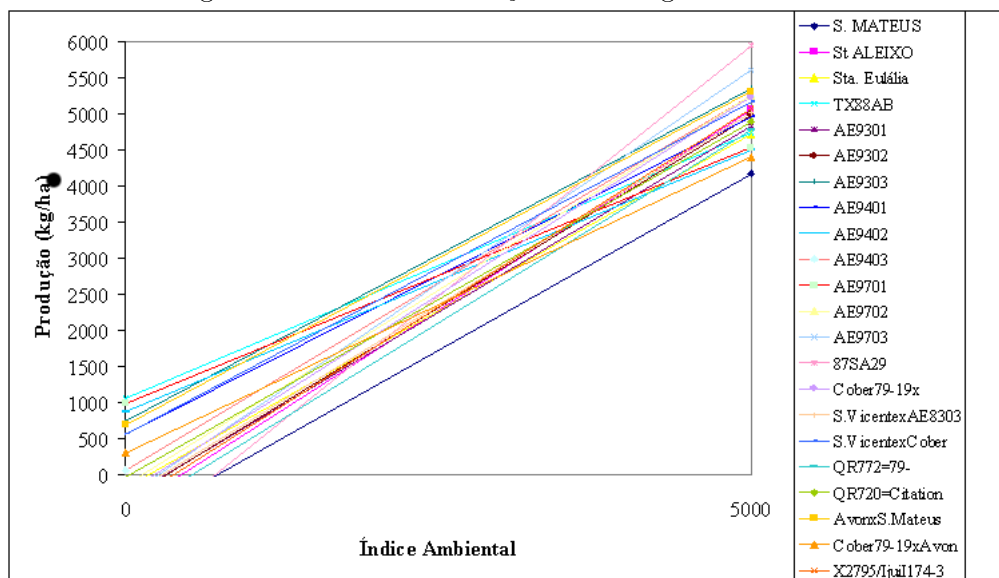


Figura B.20: Análise conjunta de regressões - 2004



Apêndice C

Programas de Simulação

Os programas de simulação foram escritos em linguagem R. Esta linguagem é escrita numa aplicação que corre em plataforma Windows a qual foi usada neste trabalho.

A página principal deste projecto que contém toda a informação disponível é <http://www.r-project.org>

C.1 Grupo 1

Número de rejeições para $\alpha = 1\%$; $\sigma_1 = \sigma_2 \in \{1, 10, 100\}$; $\mu_1, \mu_2 \in \{1, 2, 5, 7, 10, 15, 20, 25, 30\}$

```
sigma <- c(1,10,100) # Vector com os valores dos parâmetros  $\sigma_1$  e  $\sigma_2$ 
miu1 <- c(1,2,5,7,10,15,20,25,30) # Vector com os valores do parâmetro  $\mu_1$ 
miu2 <- c(1,2,5,7,10,15,20,25,30) # Vector com os valores do parâmetro  $\mu_2$ 
for (iz in sigma){
  for (ix in miu1){
    for (iy in miu2){
      # Geração de 1000 amostras obtidas a partir do produto dos valores de duas
      # variáveis com distribuição normal
      for (j in 1:1000){
        v <- c(ix,iz,iy,iz)
        # Geração de NPAs de uma variável normal de parâmetros  $\mu = 0$  e  $\sigma = 1$ 
        xx <- rnorm(200, mean=0, sd=1)
```

```

# Cálculo dos valores da variáveis  $X_1$ ,  $X_2$  e da variável produto
for (i in 1:100){
x[i]=xx[2*i-1]*v[2]+v[1]
y[i]=xx[2*i]*v[4]+v[3]
z[i]=x[i]*y[i]}
sort (z)
zz <- sort(z) # Ordenação da amostra
res <- ks.test(zz,"pnorm", mean=v[1]*v[3], sd=v[1]*v[4]) #Aplicação do teste
de Kolmogorov-Smirnov
r <- cbind("p-value=",round(res$p.value,4))
a[j] <- (round(res$p.value,4))
if (a[j]<0.01) cont1=cont1+1 # Contagem do número de casos em que a normal-
idade é rejeitada
# Escrita em ficheiro de dados
write(t(cont1),file="rejeicoestabelas.txt",append=TRUE,ncol=20)
}}}
```

C.2 Grupo 2

Número de rejeições para $\alpha = 1\%$; $\sigma_1 = \sigma_2 = 1, 100$; $\mu_1 = 1, 2, \dots, 100$; $\mu_2 = 1, 2, \dots, 20$

```

sigma <- c(1,100) # Vector com os valores dos parâmetros  $\sigma_1$  e  $\sigma_2$ 
miu1 <- seq(1,100,by=1) # Vector com os valores do parâmetro  $\mu_1$ 
miu2 <- seq(1,20,by=1) # Vector com os valores do parâmetro  $\mu_2$ 
for (iz in sigma){
for (ix in miu1){
for (iy in miu2){
# Geração de 1000 amostras obtidas a partir do produto dos valores de duas
variáveis com distribuição normal
for (j in 1:1000){
```

```

v <- c(ix,iz,iy,iz)
# Geração de NPAs de uma variável normal de parâmetros  $\mu = 0$  e  $\sigma = 1$ 
xx <- rnorm(200, mean=0, sd=1)
# Cálculo dos valores das variáveis  $X_1$ ,  $X_2$  e da variável produto
for (i in 1:100){
x[i]=xx[2*i-1]*v[2]+v[1]
y[i]=xx[2*i]*v[4]+v[3]
z[i]=x[i]*y[i]}
sort (z)
zz <- sort(z) # Ordenação da amostra
res <- ks.test(zz,"pnorm", mean=v[1]*v[3], sd=v[1]*v[4]) #Aplicação do teste
de Kolmogorov-Smirnov
r <- cbind("p-value=",round(res$p.value,4))
a[j] <- (round(res$p.value,4))
if (a[j]<0.01) cont1=cont1+1 # Contagem do número de casos em que a normal-
idade é rejeitada
# Escrita em ficheiro de dados
write(t(cont1),file="rejeicoesgraf12.txt",append=TRUE,ncol=20)
}}}
```

C.3 Grupo 3

Número de rejeições para $\alpha = 1\%$, $\sigma_1 = \sigma_2 = 1, 100$; $\mu_1 = 1 + 100 * i$; $\mu_2 = 1 + 100 * i$; $i = 0, 1, 2, \dots, 100$

```

sigma <- c(1,100) # Vector com os valores dos parâmetros  $\sigma_1$  e  $\sigma_2$ 
miu1 <- seq(1,10001,by=100) # Vector com os valores do parâmetro  $\mu_1$ 
miu2 <- seq(1,10001,by=100) # Vector com os valores do parâmetro  $\mu_2$ 
for (iz in sigma){
for (ix in miu1){
for (iy in miu2){
```

```
# Geração de 1000 amostras obtidas a partir do produto dos valores de duas
variáveis com distribuição normal
for (j in 1:1000){
v <- rnorm(2, mean=0, sd=1)
# Geração de NPAs de uma variável normal de parâmetros  $\mu = 0$  e  $\sigma = 1$ 
xx <- rnorm(200, mean=0, sd=1)
# Cálculo dos valores das variáveis  $X_1$ ,  $X_2$  e da variável produto
for (i in 1:100){
x[i]=xx[2*i-1]*v[2]+v[1]
y[i]=xx[2*i]*v[4]+v[3]
z[i]=x[i]*y[i]}
sort (z)
zz <- sort(z) # Ordenação da amostra
res <- ks.test(zz,"pnorm", mean=v[1]*v[3], sd=v[1]*v[4]) #Aplicação do teste
de Kolmogorov-Smirnov
r <- cbind("p-value=",round(res$p.value,4))
a[j] <- (round(res$p.value,4))
if (a[j]<0.01) cont1=cont1+1 # Contagem do número de casos em que a normal-
idade é rejeitada
# Escrita em ficheiro de dados
write(t(cont1),file="rejeicoesgraf34.txt",append=TRUE,ncol=20)
}}}
```

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 Sériés 9, Institute of Statistics, University of North Carolina, Chapel Hill.
- [5] Brancourt-Humel, M., Lecomte, C. (1994) - "*Selection et stabilité du rendement chez le blé tendre d'hiver*". *Agronomie*, 14 : 611-625.
- [6] Caliński, T. (1987a) - *A model for the analysis of a series of experiments repeated at several places over a period of years, I. Theory*. Biuletyn Oceny Odmian, 17-18:7-33.
- [7] Caliński, T. (1987b) - *A model for the analysis of a series of experiments repeated at several places over a period of years, II. Example*. Biuletyn Oceny Odmian, 17-18:35-71.
- [8] Caliński, T. (1997) - *Recovery of inter-block information when the experiment is in a nested block design*. *Listy Biometryczne-Biometrical Letters*, vol 34, n^o1 : 9-26.
- [9] Caliński, T., Caussinus, H. (1989) - *A note on the analysis of covariance: efficiency of concomitant variables*. *Journal of Statistic Planning and Inference*, 21 : 315-326.

- [10] Caliński, T., Czajka, S. and Kaczmarek, Z. (1997) - *A multivariate approach to analysing genotype-environment interactions. Advances in Biometrical Genetics*. Proceedings of the Tenth Meeting of the EUCARPIA. Section Biometrics in Plant Breeding, P. Krajewski and Z. Kaczmarek (eds.), 3-14.
- [11] Caliński, T., Czajka, S., Kaczmarek, Z. and Lejeune, M. (1995a) - *The application of factor analysis in interpreting results of a series of variety trials, II. Example*. Biuletyn Oceny Odmian, 26-27:23-28.
- [12] Caliński, T., Kageyama, S. (1996) - *The randomization model for experiments in block designs and the recovery of inter-block information*. Journal of Statistical Planning and Inference, 52:359-374.
- [13] Cook, R. D., Weisberg, S. (1999) - *Applied Regression Including Computing and Graphics*. John Wiley & Sons, Inc.
- [14] Currie, I. D. and Durban, M. (2000) - *Adjusted profile score: some applications*. Statistical Modelling. Proceedings of the 15th International Workshop on Statistical Modelling, Bilbao, Spain, July 17-21:404-407.
- [15] D'Agostino and Stephens (1986) - *Goodness-of-fit Techniques*. Marcel-Dekker, New York.
- [16] Dias, G. C. (1994) - *Selective F tests*. Trabalhos de Investigação, nº1. FCT-UNL
- [17] Dias, C. P. (2000) - *Análise Conjunta Pesada de Regressões*. Tese de Mestrado. Universidade de Évora.
- [18] Digby, P. G. N. (1979) - *Modified joint regression analysis for incomplete variety x environment data*. - J. Agric. Sci., Camb., 93 : 81 - 86.
- [19] Eberhart, S. A. and Russel, W. A. (1966) - *Stability parameters for comparing varieties*. Crop. Science, 6 : 36 - 40.
- [20] Everitt, Brian. S., Hothorn, T. (2006) - *A Handbook of Statistical Analysis using R*. Chapman & Hall/CRC
- [21] Faraway, J. J. (2006) - *Extending the Linear Model with R, Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC.

- [22] Ferreira, T.; Mexia, J. T.; Baeta, J.; Maçãs, B.; Gusmão, L.; Bagulho, F. e Coutinho, J. (1990) - *Comportamento relativo das cultivares de triticale 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.
- [23] Finlay, K. W. and Wilkinson, G. N. (1963) - *The analysis of adaptation in a plant-breeding programme*. - Aust. J.Aric. Res., 14 : 742 - 754.
- [24] 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.
- [25] Friedman, M. (1937) - *The use of ranks to avoid the assumption of normality implicit in the analysis of variance*. J. Am. Statistics, 11 : 86 - 92.
- [26] Friedman, M. (1940) - *A comparison of alternative tests of significance for the problem of m rankings*. Annals of Mathematical Statistics, 11 : 86 - 92.
- [27] Giesbrecht, F. G. and Gumpertz, M. L. (2004) - *Planning, Construction and Statistical Analysis of Comparative Experiments*. John Wiley & Sons, Inc.
- [28] Guerreiro, J. S. (1970) - *Curso de Matemáticas Gerais: Noções de Álgebra Linear*. Vol IV - Escolar Editora, Lisboa.
- [29] Gusmão, L. (1985) - *An adequate design for regression analysis of yield trials*. Theor. Appl. Genet., 71: 314 - 319.
- [30] Gusmão, L. (1986a) - *Inadequacy of blocking in cultivar yield trials*. Theor. Appl. Genet., 72: 98 - 104.
- [31] 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.
- [32] Gusmão, L. (1988) - *Assessing small grain cultivars for yield in variable environments through Joint Regression Analysis*. Rachis, 7: 22 - 25.
- [33] Gusmão, L. (1990) - *Avaliação de cultivares (uma perspectiva experimental diferente)*. Revista de Ciências Agrárias - n°2, vol. 13: 11 - 18.
- [34] 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.

- [35] 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.
- [36] 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.
- [37] 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.
- [38] Gusmão, L.; Fernandes, D.; Maçãs, B.; Mexia J.T.; Gusmão, M. R. (1995) - *Análise pela "RECOR" dos resultados de produção numa rede regional de ensaios de adaptação 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.
- [39] Gusmão, L.; Rebelo, D. C.; Miranda, M.J.; Baeta, J. e Vaz, F. (1993a) - *Assessment of specific genotype \times environment interactions in Lupinus luteus L.* *Brotéria Genética*, Lisboa, 14: 141 - 144.
- [40] 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.
- [41] Gray, E. (1982) - *"Genotype-environment interactions and stability analysis for forage yield of orchardgrass clones"*. *Crop Science* 22: 19 - 23.
- [42] Hedges, L.V., Olkin, I. (1985) - *Statistical Methods for Meta-Analysis*. Academic Press, London.
- [43] Ito, K. (1980) - *Robustness of Anova and Macanova test procedures*. P.R. Krishnaiah (ed) - *Handbook of Statistics - vol.II*, North Holland.
- [44] Jalalaluddin, M.D., Harrison, S.A. (1993) - *"Repeatability of stability estimators for grain yield in wheat"*. *Crop Science* 33: 720-725.
- [45] John, J.A., Williams, E.R. (1995) - *Cyclic and Computer generated designs*. Chapman and Hall, London, 2nd ed.

- [46] Judge, George G.; Hill, R. Carter; Griffiths, William E.; *Lütkepohl*, Helmut L. and Lee, Tsoung-Chao, (1985) - *The Theory and Practice of Econometrics*. John Wiley and Sons, Inc. 2nd edition.
- [47] 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.
- [48] Kang, M.S., Gauch, H.G. (1996) - *Genotype by Environmental Interaction*. CRC Press, New York.
- [49] Lin, C. S., Binns, M. R. and Lefkovich, L. P. (1986) - *Stability analysis: Where Do We Stand?* Crop. Science, 26: 894 - 900.
- [50] Loève, M. (1961) - *Probability Theory*. 2nd ed. Van Nostrand Company, New York.
- [51] Maçãs, B.; Coutinho, J.; Bagulho, F. Mexia, J. T.; Baeta, J.; Gusmão, L. e Ferreira, T. (1990) - *Comportamento relativo das cultivares de triticale produzidas na E.N.M.P., no período 1987/88 e 1988/89: I. Aferido pela análise convencional*. Em "V Reunião Portuguesa sobre Triticale- Elvas, Maio 1990 - Melhoramento, vol. 32.
- [52] Mason, R.L.; Gunst, R.F. and Hess, J.L.(2003) - *Statistical Designs and Analysis of Experiments with Application to Engineering and Science*. 2nd Ed. John Wiley & Sons Publication.
- [53] Mexia, J. T. (1987) - *Multi-treatment Regression Designs*. Trabalhos de Investigação, n° 1, 1987 - Departamento de Matemática, FCT-UNL.
- [54] 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.
- [55] 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.
- [56] Mexia, J. T.; Gusmão, L. e Baeta, J. (1990) - *Homogeneidade numa zona equipotencial de adaptação e suas implicações no delineamento de ensaios de adaptação*. Garcia de Orta, Série de Estudos Agronómicos, vol 10: 65 - 70.

- [57] 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.
- [58] Mexia, J. T. (1995) - *Introdução à Inferência Estatística Linear*. Edições Lusófonas, Lisboa.
- [59] Mexia, J. T.; Dias, G.C. and Neves, M. (1996) - *Roots of linear regressions, extensions and applications to joint regression analysis*. Trabalhos de Investigação n°1. FCT/UNL.
- [60] Mexia, J. T.; Amaro, A. P., Gusmão, L. & Baeta J. (1997) - *Upper Contour of a Joint Regression Analysis*. J. Genet. and Breed., 51 : 253 - 255.
- [61] Mexia, J. T.; Pereira, D. G. & Baeta, J. (1999) - *L₂ environmental indexes*. Biometrical Letters, 36: 137 - 143.
- [62] Mexia, J. T.; Pereira, D. G. & Baeta, J. (2001) - *Weighted linear joint regression analysis*. Biometrical Letters, 38: 33 - 40.
- [63] Mexia, J. T. & Pereira, D. G. (2001) - *Joint regression analysis for winter rye cultivars using L₂ indexes*. Colloquium Biometryczne, 31: 207 - 212.
- [64] Mexia, J. T. and Corte Real, P. (2001a) - *Extension of Kolmogorov's Strong Law to Multiple Regression*. 23rd European Meeting of statisticians, Funchal - Madeira. Revista de Estatística, vol. II, 2º Quadrimestre de 2001, pag. 277 - 278.
- [65] Mexia, J. T. and Corte Real, P. (2001b) - *Strong Law of Large Numbers for Additive Extremum Estimators*. Discussiones Mathematicae, Probability and Statistics, 21: 81 - 88.
- [66] Mexia, J. T. & Pinto, I. (2001) - *Generalized additivity and incomplete unbalanced design*.- Colloquium Biometryczne, 31: 201 - 206.
- [67] Mexia, J. T. & Pinto, I. (2002) - *Test for Linear restrictions, generalized additivity and consequences of restrictions violations*. Colloquium Biometryczne, 32: 197 - 204.
- [68] Mexia, J. T. & Pinto, I. (2003) - *Joint Regression Analysis and Connectedness - Application to Grapevines*- Colloquium Biometryczne, 33: 235 - 248.

- [69] Mexia, J. T. & Pinto, I. (2004) - *Comparison of wheat cultivars in Portugal (1986-2000)*. Colloquium Biometryczne, Tom 34.
- [70] Mickey, R.M.; Dunn, O.J. and Clark, V.A. (2004) - *Analysis of Variance and Regression*. John Wiley & Sons, Inc.
- [71] Mood, A. M.; Graybill, F. A. and Boes, D. C. (1974) - *Introduction to the Theory of Statistics*. McGraw-Hill, Singapura, 3rd Ed.
- [72] Mooers, C. A. (1921) - *The agronomic placement of varieties*.- J. Amer. Soc. Aron., 13: 337 - 352.
- [73] Ng, M. P. and Grunwald, G. K. (1997) - *Nonlinear regression analysis of the joint-regression model*. - Biometrics, 53 : 1366 - 1372.
- [74] Ng, M. P. and Williams, E. R. (2001) - *Joint regression analysis for incomplete two-way tables*. Aust. N.Z.J. Stat., 43(2) : 201 - 206.
- [75] Nguyen, H. T.; Sleper, D. A. and Hunt, K. L. (1980)- *"Genotype environment interactions and stability for herbage yield of tall fescue synthetics"*. Crop Science, 20 : 221 - 224.
- [76] Oliveira, Tiago (1990) - *Probabilidades e Estatística: Conceitos, Métodos e Aplicações*. Vol. II McGraw-Hill, Lisboa.
- [77] Oliveira, A. (1999) - *Misturas de Normais - uma Aplicação - Dissertação de Mestrado*. Universidade Nova de Lisboa.
- [78] Oliveira, T.A. (1999) - *Planeamento de Experiências, Novas Perspectivas - Tese de Doutoramento*, Universidade de Lisboa.
- [79] Patterson, H.D. and Williams, E.R. (1976) - *A new class of resolvable incomplete block designs*. Biometrika, 63: 83 - 92.
- [80] Pereira, D.G. & Mexia, J. T. (2002) - *Multiple comparison in Joint Regression Analysis with special reference to variety selection*. Scientific Papers of the Agricultural University of Poznan, Agriculture, 3 : 67 - 74.
- [81] Pereira, D.G. & Mexia, J.T. (2003a) - *The use of Joint Regression Analysis in selecting recommended cultivars*. Biuletyn Oceny Odmian (Cultivar Testing Bulletin), 31 : 19 - 25.
- [82] Pereira, D.G. & Mexia, J.T. (2003b) - *Reproductibility of Joint Regression Analysis*. Colloquium Biometryczne, 33 : 279 - 291.

- [83] Pereira, D.G. (2004) - *Análise Conjunta Pesada de Regressões em Redes de Ensaios*. - Tese de Doutoramento Universidade de Évora.
- [84] Perkins, J.M.; Jinks, J.L. (1968a) - "Environmental and genotype-environmental components of variability. III Multiple lines and crosses"; *Heredity*, 23: 339 - 356.
- [85] Pinto, I.M.S. (2006) - *Análise Conjunta de Regressões e Planos de Melhoramento*. - Tese de Doutoramento Universidade Nova de Lisboa.
- [86] Pollock, D.S.G. (1979) - *The algebra of econometrics*. John Wiley & Sons, Inc.
- [87] Raghavarao, D. (1971) - *Construction and combinatorial Problems in Design of Experiments*. John Wiley, New York. Reprinted (1988) by Dover with some addendum.
- [88] Ramos, L.(2007)- *Quase Normalidade e Inferência para Séries de Estudos Emparelhadas*. Tese de Doutoramento, Universidade Nova de Lisboa.
- [89] Romagosa, I. & Fox, P. N.(1993) - "Genotype x environmental interaction and adaptation. *En: Plant Breeding: Principles and Prospects*. Chapman and Hall, London, 373-390.
- [90] Ruppert, D.; Wand, M.P. and Carroll, R.J. (2003) - *Semiparametric Regression*. Cambridge University Press.
- [91] Scheffé, H. (1959) - *The Analysis of Variance*. John Wiley and Sons - New York.
- [92] Seber, G.A.F. (1980) - *The Linear Hypothesis: A General Theory*, 2nd ed., Charles Griffin and Co - London.
- [93] Seber, G.A.F. and Lee, J. A. (2003) - *Linear Regression Analysis*, 2nd ed., John Wiley and Sons - New York.
- [94] Severine, T. A. (2000) - *Likelihood methods in statistics*. Oxford University Press, New York.
- [95] Sheskin, David J. (2003) - *Handbook of Parametric and Nonparametric Procedures, third Edition(Hardcover)*. Chapman & Hall/CRC
- [96] Shrikhande, S.S. and Raghavarao, D. (1963) - *A method of construction of incomplete block designs*.- *Sankhya Ser.A* 25 : 399 - 402.

-
- [97] Shrikhande, S.S. and Raghavarao, D. (1964) - *Affine α -resolvable incomplete block designs*. In: C.R. Rao (ed.), *Contributions to Statistics*. Pregamon Press, Statistical Publishing Society, Calcutta, p. 471 - 480.
- [98] Silvey, S.D.(1975) - *Statistical Inference*. Chapman and Hall, Cambridge
- [99] Siegel, S. and Castellan, J. (1988) - *Nonparametric statistics for the behavioral science*.- 2nd Ed., Mc Graw-Hill, New York.
- [100] Toler, J. E. (1990) - "*Patterns of genotypic performance over environmental arrays*". Clemson, Thesis (PhD) - Clemson University.
- [101] Toler, J. E.; Burrows, P.M. (1998) - "*Genotypic performance over environmental arrays: A non-linear grouping protocol*". *Journal of Applied Statistics*, Abingdon, v25, n1, 131 - 143.
- [102] Toutenburg, H. (2002) - *Statistical Analysis of Designed Experiments*. Springer.
- [103] Westcott, B. (1986) - *Some methods of analysing genotype-environment interaction* *Heredity*, 56 : 243 - 253.
- [104] Williams, E. R. (1975) - *A new class of resolvable block designs*. Ph'D Thesis. University of Edinburgh.
- [105] Yates, F. and Cochran, W. G. (1938) - *The analysis of groups experiments*. *J. Agric. Sci., Cambridge*, 28 : 556 - 580.
- [106] Zobel *et al* (1988) - "*Statistical analysis of a yield trial*". *Agron. Journal*, 80 : 388 - 393.