

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

**“Análise Conjunta de Regressões Lineares:
Revisão e Desenvolvimentos Computacionais em R”**

Adriano Arcanjo Monteiro

Mestrado em Estatística, Matemática e Computação na área de
Especialização em Estatística Computacional

Lisboa, 2015

UNIVERSIDADE ABERTA



**“Análise Conjunta de Regressões Lineares:
Revisão e Desenvolvimentos Computacionais em R”**

Adriano Arcanjo Monteiro

Dissertação apresentada na Universidade Aberta para obtenção do Grau de
Mestre em Estatística, Matemática e Computação na especialização em
Estatística Computacional

Dissertação orientada pelo
Professor Doutor Amílcar Manuel do Rosário Oliveira

Lisboa, 2015

Dedicatória

*A Minha esposa Ana Paula Rocha Monteiro e
aos meus filhos Anderson Monteiro e Andercia Monteiro*



Resumo

O R é em simultâneo uma linguagem de programação e um ambiente para computação estatística e gráficos. Estaremos interessados em compreender a funcionalidade dos *packages* do R, criando um novo, capaz de executar automaticamente a Análise Conjunta de Regressões, ACR.

A ACR é uma técnica muito utilizada para interpretar redes de ensaios no domínio da avaliação de cultivares. Nesta técnica encontramos um conjunto de procedimentos que nos permite na sua utilização fazer a comparação e seleção de cultivares. Numa primeira fase recorreremos ao algoritmo zigzag no ajustamento de regressões lineares, uma por cultivar, onde nas regressões a variável controlada é o índice ambiental que mede a produtividade dos vários ambientes. Os valores dos índices ambientais e dos coeficientes das regressões serão ajustados simultaneamente pelo referido algoritmo, que segundo Oliveira, A. (2007) se baseia num processo iterativo onde as minimizações alternam entre os coeficientes de regressão e os índices ambientais.

Para avaliarmos a qualidade do ajustamento obtida pelo algoritmo calcularemos o coeficiente de determinação R^2 através dum modelo linear para os resíduos.

Contrariamente ao que é usual, neste caso estamos interessados em obter valores baixos de R^2 , pois um valor próximo a zero indica um bom ajuste, enquanto valores maiores indicam que o algoritmo zigzag não foi capaz de extrair toda a informação relevante dos dados.

Através do Método do Contorno Superior exemplificaremos, quais os cultivares integram esse contorno e posteriormente, através do teste t e dos métodos de comparação múltipla de Scheffé, Bonferroni e Tukey, averiguaremos quais os cultivares, ao nível de significância de 5%, são significativamente dominantes ou dominados, nos determinados intervalos de dominância.

Neste trabalho, após apresentação dos aspetos teóricos, no Capítulo 5 desenvolveremos um conjunto de funções e rotinas, criando um “*package*” em linguagem R que nos permitirá, analisar no Capítulo 6, um conjunto de dados simulados para verificação da aplicabilidade do mesmo.

Através da utilização de dados simulados referentes ao ensaio de cultivares, mostraremos como utilizar a ACR em planos de melhoramento com auxílio ao *package* 'jra', desenvolvido para cumprir um dos objetivos desta dissertação.

Palavras chave: Linguagem R; Análise Conjunta de Regressões; Método do contorno superior; Teste t; Métodos de comparação múltipla de Scheffé, Bonferroni e Tukey.

Abstract

R is simultaneously a programming language and an environment for statistical computation and graphics. We will be interested in understanding the functionality of the packages of R, creating a new one, capable to execute the Joint Regression Analysis, ACR, automatically.

The JRA is a technique very used to interpret networks in the domain of the evaluation of cultivars. In this technique we find a group of procedures that allow us to do the comparison and selection of cultivars. In a first phase we resorted upon the zigzag algorithm in the adjustment of linear regressions, one for cultivar, where in the regressions the controlled variable is the environmental index that measures the productivity of the several environments. The values of the environmental indexes and of the coefficients of the regressions, will be adjusted simultaneously by the referred algorithm, that according to Oliveira, A. (2007) is based on a iterative process where the minimizations alternate between the regression coefficients and the environmental indexes.

In order to evaluate the quality of the adjustment obtained by the algorithm we will calculate the determination coefficient R^2 using a linear model for the residues.

Contrarily to what is usual in this case we are interested in obtaining low values of R^2 , because a close value to zero indicates a good adjustment, while larger value indicate that the zigzag algorithm was not capable to extract all the relevant information from the data.

Through the Upper Contour Method, we will exemplify, which cultivars integrate that Contour and later, observed through t test and the methods of multiple comparisons of Scheffé, Bonferroni and Tukey, which cultivars, at level of significance of 5%, are significantly dominant or dominated, in the determined intervals of dominance.

In this work, after presenting the theoretical aspects, in Chapter 5 we will develop a set of functions and routines, creating a "package" in language R that will allow us, to analyze in Chapter 6, a set of simulated data for verification of the applicability of the same.

Through the use of simulated data to test cultivars, we will show how to use JRA in improvement plans with the aid of the package '**jra**', developed to accomplish one of the main objectives of this dissertation.

key Words: R Language; Joint Regression Analysis; Upper Contour Method; t test; Methods of multiple comparison of Scheffé, Bonferroni and Tukey.

Agradecimentos

Para elaboração deste trabalho, tive a colaboração direta ou indireta de várias pessoas que muito contribuíram para tornar esta tarefa menos difícil e mais agradável. Por isso dedico aqui os meus agradecimentos.

Ao Professor Doutor Amílcar Manuel do Rosário Oliveira, meu Orientador, a quem reconheço pelo grande empenho na orientação durante a elaboração deste trabalho.

Aos professores da parte curricular do mestrado pela disponibilidade e vontade de partilharem os seus conhecimentos e as suas experiências.

À minha esposa pelo apoio incondicional e a sua paciência em todos os momentos ao longo os anos de mestrado. Sempre me apoiou quando mais necessitei.

Aos filhos Anderson Monteiro e Andercia Monteiro pela compreensão que tiveram pela minha ausência em diversos momentos.

Ao meu amigo André Machado de Oliveira pela troca e partilha de conhecimentos relativos aos desenvolvimentos computacionais no R.

A todos os colegas pela oportunidade de estudarmos juntos Estatística Computacional e em particular Rute Martins, agradeço pelo constante incentivo na realização de todos os trabalhos que nos foram propostos.

Agradeço também a todos que de alguma forma colaboraram com este trabalho.

Conteúdo

1. Introdução.....	3
2. A Linguagem de Programação R	9
2.1. Definição e Utilidade	9
2.2. Breve historial da linguagem R.....	10
2.3. O Ambiente R	11
2.3.1. Entrada de dados no R.....	15
2.3.2. Ajuste do modelo linear.....	16
2.4. Programação no R.....	19
2.4.1. Estruturas de Controlo de Programação	20
2.4.2. Funções definidas pelo utilizador.....	21
3. Análise Conjunta de Regressões Lineares.....	25
3.1. Nota histórica da ACR	25
3.2. Aspetos preliminares	27
3.2.1. Teorema de Scheffé.....	27
3.2.2. Ajustamento aos Mínimos Quadrados Estruturados	30
3.3. Modelos para blocos incompletos.....	32
3.3.1. Função objetivo	32
3.3.2 Algoritmo zigzag.....	35
3.3.3 Análise dos Resíduos.....	39
3.4. Modelos para blocos completos	41
3.4.1. Técnica de Dupla Minimização.....	42
3.4.2. Testes de razão de verosimilhanças.....	45
3.4.3. Análise de variância	51

4. Método do Contorno Superior	57
4.1. Seleção pelo método do Contorno Superior	57
4.1.1. Geometria do Contorno Superior.....	57
4.1.2. Métodos de comparação de cultivares	67
5. Package em linguagem R para a execução automática da técnica ACR.	75
5.1. Documentação do package	75
5.2. Estrutura do <i>package</i>	81
6. Aplicação prática e Resultados computacionais	95
6.1. Aplicação prática.....	95
6.1.1. Algoritmo zigzag	96
6.1.2. Análise do contorno superior.....	100
6.1.3. Comparações múltiplas ao nível de significância 5%	101
6.2. Resultados computacionais	104
6.2.1. Ano de 2010.....	105
6.2.2. Ano de 2011.....	111
6.2.3. Ano de 2012.....	117
7. Conclusões e trabalhos futuros	123
Referências Bibliográficas	125
Apêndice A: Programa que permite aplicar exemplos do Método do contorno superior para os gráficos 4.1 à 4.6.....	135
Apêndice B: Tabelas de produção de cultivares (kg/hectare)	143
Apêndice C: Média dos índices e R^2 dos anos de 2010 a 2012 nos respectivos locais....	147

Índice de Tabelas

Tabela 3.1: Quadro resumo da ANOVA	54
Tabela 6.1: Lista de cultivares ensaiados.....	95
Tabela 6.2: Presença e ausência de cultivares nos vários anos.....	96
Tabela 6.3: Coeficientes de Regressão ajustados e os valores de R^2	97
Tabela 6.4: Índices ambientais e média anual.	98
Tabela 6.5: Somas de quadrados dos resíduos e coeficiente R^2	99
Tabela 6.6: Médias dos índices ambientais e coeficientes R^2 por bloco	99
Tabela 6.7: Composição dos contornos superiores ano a ano	100
Tabela 6.8: Cultivares dominantes e dominados: resultados da aplicação do teste t de Student ao nível de significância de 5%	101
Tabela 6.9: Cultivares dominantes e dominados: resultados da aplicação do método de Scheffé ao nível de significância de 5%.	102
Tabela 6.10: Cultivares dominantes e significativamente dominados: resultados da aplicação do método de Bonferroni ao nível de significância de 5%	103
Tabela 6.11: Cultivares dominantes e significativamente dominados: resultados da aplicação do método de Tukey ao nível de significância de 5%	104
Tabela B.1: Produção de cultivares (kg/ha) do ano de 2010.....	143
Tabela B.2: Produção de cultivares (kg/ha) do ano de 2011	143
Tabela B.3: Produção de cultivares (kg/ha) do ano de 2012	144

Índice de Figuras

Figura 2.1: Regressão linear simples	18
Figura 4.1: Interseção das retas ajustadas	59
Figura 4.2: Retas paralelas.....	60
Figura 4.3: Traçado do contorno superior quando $\theta_1 \leq \theta_{min}$	62
Figura 4.4: Traçado do contorno superior para $\theta_{min} < \theta_1 < \theta_{max}$	63
Figura 4.5: Traçado do contorno superior para $\theta_1 \geq \theta_{max}$	65
Figura 4.6: Contorno superior com dois cultivares dominantes	67
Figura 6.1: Traçado do contorno superior – 2010	106
Figura 6.2: Traçado do contorno superior – 2011	112
Figura 6.3: Traçado do contorno superior – 2012	118

Lista de símbolos e siglas

ACR	Análise Conjunta de Regressões.
zigzag	algoritmo ziguezague.
CRAN	Comprehensive R Archive Network.
AT&T	Laboratórios Lucent Technologies.
$\text{car}(V)$	característica da matriz V .
β	coeficiente angular de regressão linear.
R^2	coeficiente de determinação.
Ω^\perp	complemento ortogonal do espaço Ω .
\aleph	define a família das matrizes do modelo.
$(\tilde{\alpha}_{j,l}, \tilde{\beta}_{j,l})$	coeficientes de regressão ajustados para o j -ésimo cultivar no l -ésimo ano.
$\Omega^0 = R[X(\theta_b^0)],$	espaço imagem da matriz $X(\theta_b^0)$.
$\mathbb{R}_b \times \mathbb{R}_j \times \mathbb{R}_l$	espaço obtido pelo produto cartesiano.
Γ	espaço paramétrico dos vetores de estrutura.
Ω	espaço paramétrico total.
$(\hat{\alpha}_j, \hat{\beta}_j)$	estimadores de máxima verosimilhança para α_j e β_j .
$\hat{\sigma}^2$	estimador de máxima verosimilhança para σ^2
$(\tilde{\alpha}_j^1, \tilde{\beta}_j^1)$	estimadores dos mínimos quadrados para (α_j, β_j) obtidos na primeira iteração do algoritmo zigzag.
$S_1(X^b)$	função a minimizar em ordem a X^b na primeira iteração do algoritmo zigzag.
$S_1(\alpha^J, \beta^J)$	função objetivo a minimizar na primeira iteração do algoritmo

	zigzag.
$S(\omega_k \theta_{1,b})$	função objetivo a minimizar em ordem a ω_k dado $\theta_{1,b}$.
GPL	General public license.
x_i	índice ambiental do i-ésimo bloco, par (local, ano).
\tilde{x}_i	índice ambiental ajustado do i-ésimo bloco.
L_2	índice ambiental L_2 .
$\ln(L(\alpha^J, \beta^J, X^b, \sigma^2))$	logaritmo da função de máxima verosimilhança.
$\mathbf{X}(\theta_b)$	matriz do modelo que depende do vetor de estrutura θ_b .
Hess(S)	matriz hessiana de S.
$(\triangleleft v_b, \triangleright v_b)$	mínimo e máximo das componentes do vetor v_b .
$n \times k$	n linhas e k colunas numa matriz.
b	número de blocos.
J	número de cultivares estudados.
(α_j, β_j)	ordenada na origem e coeficiente angular (ou declive) da reta de regressão a ajustar para o j-ésimo cultivar
$p_{i,j}$	pesos atribuído ao j-ésimo cultivar $j = 1, \dots, J$, no i-ésimo bloco, $i = 1, \dots, b$.
$\theta_{j,j'}$	ponto de intersecção entre as retas de regressão ajustadas com índices j e j' .
\otimes	produto de kronecker.
$t_{1-\alpha, k, g}$	quantis da distribuição t de student com k e g graus de liberdade ao nível α de significância.
ζ	relação de equivalência entre matrizes do modelo.
$Y_{i,j}$	rendimento do j-ésimo cultivar no i-ésimo bloco.
$r_{i,j}$	resíduo associado ao j-ésimo cultivar no i-ésimo bloco.

StatLib	sistema de distribuição de <i>softwares</i> estatísticos
\tilde{S}	soma das somas dos quadrados dos resíduos das regressões, correspondentes à última iteração do algoritmo zigzag.
\oplus	soma direta ortogonal de matrizes.
\dot{X}^b	transposta de X^b .
\dot{Y}	transposta de Y .
α	valor comum das ordenadas na origem.
$F_{1-\alpha,k,g}$	quantis da distribuição F de Snedecor central com k e g graus de liberdade ao nível α de significância.
$t_{j,j'}(\theta_0)$	valor da estatística do teste t para comparação dos cultivares j e j' .
$E(Y_n)$	valor esperado do vetor Y_n .
x_i^0 ,	valor inicial para o índice ambiental do i -ésimo bloco.
(α^J, β^J)	vetor das ordenadas na origem e dos declives para as regressões.
$\alpha 1^J$	vetor de dimensão J com todas as componentes iguais a α .
$\beta 1^J$	vetor de dimensão J com todas as componentes iguais a β .
X^b e θ_b	vetor de estrutura cujas componentes são os b índices ambientais.
Y_n	vetor de observações da variável dependente com n componentes.
$\varepsilon_{i,j}$	vetor dos erros no modelo para os índices ambientais.
\tilde{X}^b e $\tilde{\theta}_b$	vetor dos índices ambientais ajustados.
\tilde{X}_1^b e $\tilde{\theta}_{1,b}$	vetor estimado para X^b e θ_b obtido na primeira iteração.

CAPÍTULO 1
INTRODUÇÃO

1. Introdução

Em estatística encontramos as mais diversas técnicas de análise de dados que nos permitem obter as melhores conclusões em cada problema, contudo nem sempre a melhor técnica é usada.

A técnica que vamos apresentar no nosso trabalho é a Análise Conjunta de Regressões, ACR, pois pertence ao grupo das técnicas baseadas no ajustamento de regressões e é utilizada para interpretar redes de ensaios na comparação e seleção de cultivares. Ao proceder a essa técnica uma regressão linear é ajustada por cultivar, tendo como variável dependente o rendimento. A variável controlada, o chamado índice ambiental, mede, para cada par (local, ano), a respetiva capacidade produtiva. Assim o coeficiente angular β medirá a taxa de aumento da produção com a melhoria das capacidades.

Utilizamos o algoritmo zigzag no ajustamento de regressões lineares, pois baseia-se num processo iterativo onde as minimizações alternam entre os coeficientes de regressão e os regressores não observáveis (índices ambientais) que estão relacionados a uma variedade de características que não podem ser diretamente aferidas, o que se torna difícil identificar.

A utilização do método do contorno superior na comparação e seleção de cultivares foi introduzida por Mexia et al. (1997). Podemos verificar que o contorno superior é uma poligonal convexa pois a mesma é constituída por segmentos pertencentes às regressões dominantes.

De acordo com Pinto (2006) e Oliveira (2007), a informação fornecida pelo traçado do contorno superior permite facilmente identificar um conjunto de cultivares que se destacam dos restantes para certos valores do índice ambiental, isto é quando duas retas se interseitam num determinado ponto (ponto de interseção), à direita do mesmo ponto, a reta dominante é a que possui maior declive e à esquerda, do mesmo, a reta dominante é a que possui menor declive. Assim devemos selecionar o contorno superior resultante da interseção das retas de regressão linear associadas aos cultivares em estudo.

O objetivo principal desta dissertação consiste, em sintetizar num só trabalho a revisão dos principais fundamentos da técnica ACR e a criação de um conjunto de procedimentos computacionais em linguagem R que permitirá a execução automática dessa

CAPÍTULO 1

INTRODUÇÃO

técnica, nomeadamente através das implementações do Algoritmo zigzag, do Método do contorno superior e Comparações múltiplas de Scheffé, Bonferroni e Tukey.

Na execução automática da ACR pretendemos apresentar casos práticos que ilustrem a sua aplicação em modelos para blocos completos, isto é casos em que estão presentes todos os cultivares em todos os blocos, onde o rendimento médio por bloco é um bom valor inicial para a busca dos índices.

Esta dissertação, para além do primeiro capítulo correspondente à Introdução e do sétimo capítulo correspondente às Conclusões Finais e Trabalhos Futuros, encontra-se dividida em cinco capítulos fundamentais:

No *segundo capítulo*- A Linguagem de Programação R, onde, para além da definição, procuramos compreender a origem e evolução do R, mas também as funcionalidades das funções para o funcionamento dos *packages*.

O projeto R inicialmente escrito pelos professores Robert Gentleman e Ross Ihaka foi iniciado em 1996 e que é uma ferramenta bastante poderosa com boas capacidades ao nível da programação e um conjunto bastante vasto (e em constante crescimento) de *packages* que acrescentam bastantes potencialidades à já poderosa versão base do R, apesar do seu carácter gratuito.

Além de ser um programa para análises estatísticas, o R é acima de tudo uma linguagem de programação, com a qual podemos programar as nossas próprias funções.

No *terceiro capítulo* - Análise Conjunta de Regressões é feita uma revisão da literatura considerada relevante para discutirmos aspetos teóricos, focalizando os aspetos mais importantes desta técnica, isto é começamos por estudar as questões relativas ao ajustamento e a finalizar este capítulo, apresentamos os testes de razão de verosimilhanças.

No *quarto capítulo* - é dado especial destaque à apresentação do Método do Contorno Superior. Os cultivares cujas retas de regressão ajustadas integram o contorno superior, são cultivares dominantes num intervalo de dominância. Os restantes cultivares, devem ser comparados com os cultivares dominantes (pelos métodos de comparação múltipla), afim de averiguar se são significativamente dominados.

No *quinto capítulo* - desenvolveremos o algoritmo tendo em vista a criação do package “**jra**” que nos permitirá, analisar no sexto capítulo, um conjunto de dados simulados para

verificação da aplicabilidade do mesmo recorrendo ao *software* “*open source*” R, bastante versátil, conforme descrito de forma sucinta no segundo capítulo.

No *sexto capítulo* - Aplicação prática e resultados computacionais, trataremos dados simulados referentes a cultivares dos anos fictícios de 2010 a 2012.

CAPÍTULO 2
A LINGUAGEM DE PROGRAMAÇÃO R

2. A Linguagem de Programação R

2.1. Definição e Utilidade

A linguagem de programação R é o pacote de *software* livre, e de acesso aberto para gráficos científicos e para análises estatísticas, portanto, um ambiente de programação de linha de comando, significando que o utilizador não encontrará uma interface gráfica: a operação do ambiente é basicamente através de códigos digitados a partir da linha de comando do programa, sendo este um aspeto que o diferencia de grande parte do restante *software* estatístico existente.

Para os estatísticos, o R é especialmente útil porque contém diversos mecanismos incorporados para a organização de dados, execução de cálculos sobre informações e criação de representações gráficas de conjuntos de dados.

No entanto não se trata meramente de uma linguagem equiparável às restantes, visto ser simples, versátil, de fácil apreensão resolvendo facilmente os problemas da praxis do dia-a-dia de um investigador reflexivo.

R é uma implementação “*open source*” da linguagem S e funciona em várias plataformas de computação e pode ser instalado gratuitamente a partir do sítio do projeto, em <http://www.r-project.org/>, sob os termos da GNU *General Public License (GPL)*, assim como podem ser obtidos detalhes sobre todo o projeto, informações diversas e documentação de apoio. Na instalação do R é instalado o RGui, uma interface de linha de comando que em qualquer sistema operacional é extremamente simples.

Apesar do seu carácter gratuito, o R é uma ferramenta bastante poderosa com boas capacidades ao nível da programação e um conjunto bastante vasto (e em constante crescimento) de *packages* que acrescentam bastantes potencialidades à já poderosa versão base do R (Torgo, 2006). Estes *packages*, que são bibliotecas com utilização de funções específicas ou em áreas de estudo específicas, vão sendo criados e implementados à medida que os investigadores em qualquer parte do mundo vão necessitando de desenvolver aplicações nas mais diversas áreas, permitindo assim o desenvolvimento do R, sem a intervenção direta do grupo *R Core Team*.

Nesta dissertação é nosso objetivo criar desenvolvimentos no sentido da criação de um *package* que facilitará a aplicação da técnica ACR.

2.2. Breve historial da linguagem R

O atual R é o resultado de um esforço de colaboração com as contribuições de todo o mundo (Doação de códigos, correções de erros e documentação).

O sistema foi desenvolvido a partir da linguagem S, tendo as suas origens no *Bell Laboratories* (anteriormente *AT&T*, agora *Lucent Technologies*) em finais dos anos 80 por John Chambers e colegas.

Mais tarde, em 1996, foi iniciado o projeto R inicialmente escrito pelos professores Robert Gentleman e Ross Ihaka, também conhecido como "R & R" do Departamento de Estatística da Universidade de *Auckland*, Nova Zelândia, tendo em vista o desenvolvimento de um projeto estatístico com grande potencial para o público em geral. Segundo afirmam a ideia de desenvolver algo como o R surgiu de conversas que tinham nos corredores da universidade.

Desde meados de 1997, há um grupo central de estatísticos (*R Core Team*, inicialmente com 9 membros, um número que tem crescido posteriormente a “ ver no R a linha de comando `contributors()`”) com acesso de gravação para a fonte R, no entanto, esta linguagem começou a ser desenvolvida pelos mesmos professores, tendo começado a criação de *software* sem qualquer experiência em computação, como um desafio. No início dos anos 90, mais propriamente em 1991, passaram a dedicar muito tempo ao R, cerca de 5 a 6 anos a desenvolver a mesma linguagem, conforme afirma Gentleman ao jornal *New York Times* (Vance, 2009). “Por cinco ou seis anos, trabalhávamos juntos o tempo todo”. “Um digitava e o outro pensava”.

O mesmo *New York Times* confirma que alguns estatísticos que estudaram as primeiras versões do *software* consideraram que havia muito a refinar, mas a despeito dessas deficiências o R imediatamente conquistou adeptos entre as pessoas que viam as oportunidades, que o *software* livre personalizado poderia oferecer. Por isso Ross Ihaka afirmou que "A aplicação R é uma demonstração real do poder da colaboração, e não creio

que fosse possível criar algo parecido de qualquer outra maneira”. “Se tivéssemos escolhido lançar o *software* como produto comercial, teríamos vendido cinco cópias”.

A primeira distribuição do R foi disponibilizada em 1993, quando algumas cópias foram disponibilizadas no StatLib, um sistema de distribuição de *softwares* estatísticos, pois os dois professores desejavam uma tecnologia mais adaptada às necessidades dos seus alunos de estatística, que precisavam de analisar dados e produzir modelos gráficos dessas informações. Outra razão que os levou a criar a sua própria linguagem estava ligada com a complexidade dos *softwares* estatísticos que estavam disponíveis, sendo que os mesmos tinham sido criados por programadores da computação que não tinham tanta sensibilidade para a área de estatística.

Fox, John (2009) escreve ao R Journal, “embora eu não tenha formalmente pesquisado os muitos projetos de código aberto existentes, o projeto R é aparentemente incomum na sua organização formal de voluntários independentes”. “Esta estrutura tem atraído um grupo de indivíduos extraordinariamente talentosos e competentes”.

2.3. O Ambiente R

O termo “ambiente” pretende caracterizar o R como um sistema totalmente completo e coerente, em vez de uma aglomeração de ferramentas muito específicas e inflexíveis, como é o caso com outros *softwares* de análise de dados (Venables et al. 2015).

Após a instalação do R a partir do sítio do projeto basta usar ou o ícone que normalmente está disponível no desktop do Windows, ou então usar o respetivo item no menu “iniciar → Programas → R” e aparece uma janela que apresenta o *prompt* do R (`>`), com o cursor à sua frente. Torgo (2006), apresenta um pequeno exemplo do tipo de interação que vamos ter com o R, escrevendo o seguinte comando no *prompt* e carregando em seguida na tecla ENTER (esta é a forma de mandar o R executar o comando que acabamos de escrever),

```
> R.version
platform      i386-w64-mingw32
arch          i386
os            mingw32
system        i386, mingw32
status
major         3
```

CAPÍTULO 2

A LINGUAGEM DE PROGRAMAÇÃO R

```
minor          2.2
year          2015
month         08
day           14
svn rev       69053
language      R
version.string R version 3.2.2 (2015-08-14)
nickname      Fire Safety
```

O comando *R.version* ao ser executado fez aparecer uma série de informação sobre a versão do R bem como sobre o computador onde o R está a ser executado.

O R é uma linguagem orientada para objetos. Isto quer dizer que os dados manipulados são armazenados na memória ativa do computador na forma de objetos, que têm um nome e aos quais se podem aplicar ações.

Os tipos básicos de objetos do R são:

- vetores
- matrizes e arrays
- data-frames
- listas
- funções

Os quatro primeiros tipos são objetos que armazenam dados e que diferem entre si na forma da armazenar e operar com os dados. O último (funções) é um tipo objeto especial que recebe algum "*input*" e produz um "*output*". (Júnior, 2005).

Assim se por exemplo quisermos criar o objeto x com o valor 5 devemos fazer:

```
> x <- 5
```

De seguida podemos chamar o conteúdo do objeto x criando fazendo:

```
> x
```

e carregando em <ENTER> obtemos neste caso

```
[1] 5
```

O dígito 1 entre parênteses retos, indica que o conteúdo exibido inicia-se com o primeiro elemento do objeto x. Podemos armazenar um valor em um objeto com certo nome usando o símbolo < seguido de um sinal -.Exemplos:

```
> x <- sqrt(2) # armazena a raiz quadrada de 2 em x
> x           # digite o nome do objeto para ver seu conteúdo
[1] 1.414214
```

Podemos ver quais os objetos atualmente na memória do computador usando as funções `ls()` ou `objects()`. Se já não necessitamos de algum dos objetos podemos então apagá-los com a função `rm()`.

Venables, et al. (2015) afirmam que todos os objetos criados durante uma sessão de R podem ser armazenados permanentemente num arquivo para uso em futuras sessões R e que no final de cada sessão é-lhe dada a oportunidade de salvar todos os objetos disponíveis atualmente, para isso basta executar o seguinte `> q()`. Se indicarmos sim, os objetos são gravados num arquivo chamado `.RData` do diretório atual, e as linhas de comando usadas na sessão são salvas num arquivo chamado `.Rhistory`.

Gravar a sessão só terá interesse se de facto pretendermos continuar o que estávamos a fazer mais tarde.

Os ficheiros com o estado da sessão são sempre gravados no diretório atual onde o R está a funcionar. Para saber do diretório atual do R basta fazer no *prompt*,

```
> getwd()
[1] "C:/ Documents and Settings /Adri/ Os meus documentos"
```

Conhecido o diretório do R, pode ser alterado fazendo por exemplo:

```
> setwd ("C:/ Documents and Settings /Adri/ Os meus
documentos")
```

Para guardar os arquivos num local específico, deve seguir o seguinte procedimento:

1. Na barra de menus do R clique em *File* e em seguida *Change dir*
2. Na janela *Choose directory* clique em *Browse*
3. Abrir-se-á uma nova janela *Choose a folder*. Aqui deverá escolher o destino onde pretende guardar os arquivos (as diretorias estarão todas visíveis).

Everitt e Hothorn (2006) afirmam que o utilizador pode também aceder à ajuda disponível no mesmo programa, bastando escrever `help()`. Através deste comando irão aparecer

CAPÍTULO 2

A LINGUAGEM DE PROGRAMAÇÃO R

todas as informações que o programa dispõe, pois entre parênteses não colocamos nenhum argumento ou fazer uma pesquisa mais objetiva, como por exemplo `help(mean)` ou abreviado `?mean` queira exibir a documentação sobre a função média. Para um recurso especificado por caracteres especiais, o argumento deve ser colocado entre aspas duplas ou simples, tornando-se uma "cadeia de caracteres": Isso também é necessário para algumas palavras com significado sintático incluindo *if*, *for* e *function*.

```
>help("[[")
```

```
>help("function")
```

O comando `help.search` (alternatively `??`) permite-nos procurar ajuda de várias maneiras.

Nas versões Windows do R, a forma mais fácil de obter ajuda é através da utilização do menu *help* disponível na janela da aplicação R.

Packages em R são bibliotecas contendo funções e dados, que não tem uma utilidade geral, mas são importantes para alguma finalidade específica.

Nesta dissertação, pretendemos desenvolver alguns procedimentos tendo em vista a criação de um *package*, entretanto existem já milhares que podem ser usados, mas para isso necessitamos de realizar o *download* dos mesmos e depois instalá-los dentro do *software* R através do comando

```
>install.packages (nome do package),
```

ou em alternativa podemos ir ao menu *packages* e seguidamente ir à opção “*install package (s)*” onde (se o seu computador estiver ligado à internet) surgirão diversos *packages* disponíveis ver Daniel e Kamala (2009). Uma vez instalado, um *package* pode ser carregado com `library()` e o conjunto de dados acedidos como `data()`, como acontece com os *packages* internos (disponíveis no sistema), por exemplo,

```
> library(MASS)
```

Esta instrução, faz com que a partir de agora passem a estar disponíveis para utilização, todos os objetos (funções, dados, etc.) definidos nesse *package*, ver Verzani (2005). Cada *package* instalado tem uma ajuda específica que pode ser obtida no sistema de ajuda HTML do R conforme descrito acima em “*help*”.

2.3.1. Entrada de dados no R

Pode-se entrar com dados no R de diferentes formas. O formato mais adequado vai depender do tamanho do conjunto de dados, e se os dados já existem em outro formato para serem importados ou se serão digitados diretamente no R (Júnior, 2005).

Podemos entrar com dados definindo vetores com o comando `c()`, ou usando funções que criam vetores. Vejamos o exemplo.

```
> x <- c(1.54, 1.72, 1.67, 1.65, 1.59)
> x
[1] 1.54 1.72 1.67 1.65 1.59
```

Esta forma de entrada de dados é conveniente quando se tem um pequeno número de dados.

Quando os dados tem algum "padrão" tal como elementos repetidos, números sequenciais pode-se usar mecanismos do R para facilitar a entrada dos dados como vetores. exemplos.

```
> vetor1<- seq(from=5,to=10); vetor2<- 5:10; vetor3<- (1:5) *
10;
```

Podemos usar também a função `scan()`, pois este lê dados diretamente a partir da linha de comando, isto é, coloca o R em modo *prompt* onde o utilizador deve digitar cada dado seguido da tecla <ENTER>.

Para criar o objeto matriz no R vamos ter que usar a função

```
> matrix().
```

Como no exemplo seguinte:

```
> x<- matrix(c(0.2,0.1,0.5,0.4,0.9,0.8,0.4,0.3,0.5,0.7,0.1,
0.2), ncol=4)
```

Uma das formas de fazer proceder à importação dos dados é através da utilização de ficheiros em formato de texto (arquivo do tipo ASCII). Se os dados estiverem disponíveis por exemplo, numa folha do EXCEL, podemos importá-los diretamente para o R, ou aí escolher a opção <SALVAR COMO> e gravar os dados num ficheiro em formato de texto.

CAPÍTULO 2

A LINGUAGEM DE PROGRAMAÇÃO R

No R, usa-se a função `scan()` mencionada anteriormente, ou então a função mais flexível `read.table()` para importar e ler os dados, e armazenar no formato de uma *data.frame*.

Considerando que o ficheiro se encontra na diretoria C:, para importar um ficheiro de dados *cultivares.txt*, teríamos que escrever no R

```
> dados<-read.table("C:\\cultivares.txt", head=T)
```

Quando não nos lembramos da localização do ficheiro de dados que se deseja importar, a função `file.choose()` é a opção para procurar esses dados no computador, devendo proceder-se da seguinte forma:

```
> dados<-read.table (file = file.choose(), head=T)
```

O R também tem a capacidade de escolher arquivos diretamente a partir da Internet utilizando a função `url()`. Em seguida, lê página da web como se fosse um arquivo local.

```
> dados<-read.table (file=url(site), head=T)
```

O argumento *head=T*, ira indicar o nome das variáveis na primeira linha.

2.3.2. Ajuste do modelo linear

No programa R, adota-se uma abordagem unificada dirigida à modelação, o que inclui a sintaxe para especificação de modelos. Variáveis respostas e covariáveis (variáveis explicativas) são sempre especificadas da mesma forma básica, ou seja, na forma resposta~covariável, onde:

- à esquerda indica-se a(s) variável(eis) resposta
- o símbolo ~ significa é modelado por
- à direita indica-se a(s) covariável(eis)

Vamos à partida considerar um exemplo onde há apenas uma variável resposta que poderá ser explicada por uma ou mais covariáveis.

Consideremos, para o conjunto de dados existentes no R denominado de `mtcars`, o ajustamento de um modelo que explique o rendimento (Y: mpg) pelo peso do veículo (X: wt). O modelo linear é dado por:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Podendo ser ajustado no R usando a função `lm()` (*lm* : *linear model*) com o formato `lm(y~x)`. Na sintaxe da chamada função **mpg~wt** lê-se: mpg é modelado por wt, através de um modelo linear `lm()`, conforme o modelo acima. Assim, como a maioria das funções do R, podemos armazenar os resultados devolvidos pela função `lm()` num objeto.

<pre>> dados=mtcars > x=dados\$wt#peso do veiculo > y=dados\$mpg #rendimento > valores <- lm(y ~ x) > valores Call: lm(formula = y ~ x) Coefficients: (Intercept) x 37.285 -5.344</pre>	<pre>> # ou fazendo: > x=wt > y=mpg > valores<-lm(y~x,dados) > valores Call: lm(formula = y ~ x, data = dados) Coefficients: (Intercept) x 37.285 -5.344</pre>
--	---

Também podemos ter acesso aos coeficientes diretamente usando a função `coef()`. Os coeficientes do exemplo referido acima podem ser obtidos usando

```
> coef(valores)
```

Notemos que a fórmula apenas especifica a relação entre as variáveis resposta e explicativas e não implica que o modelo seja necessariamente linear. A linearidade do modelo é assumida quando usamos a função `lm()`.

Assim, podemos armazenar os resultados devolvidos pela função `lm()` num objeto. O valor devolvido pela instrução `lm()` é uma lista:

```
> valores<-lm(y~x,dados)
> is.list(valores)
[1] TRUE
```

```
> names(valores)
[1] "coefficients" "residuals" "effects" "rank" "fitted.values"
" "assign"
 [7] "qr" "df.residual" "xlevels" "call" "terms"
"model"
```

Todos os resultados apresentados são importantes para o modelo ajustado. O primeiro, e talvez o mais importante a ser examinado, é o elemento “*coefficients*” onde podemos obter um diagrama de dispersão fazendo:

```
> plot(x, y) ou de forma direta with(dados, plot(y ~ x))
```

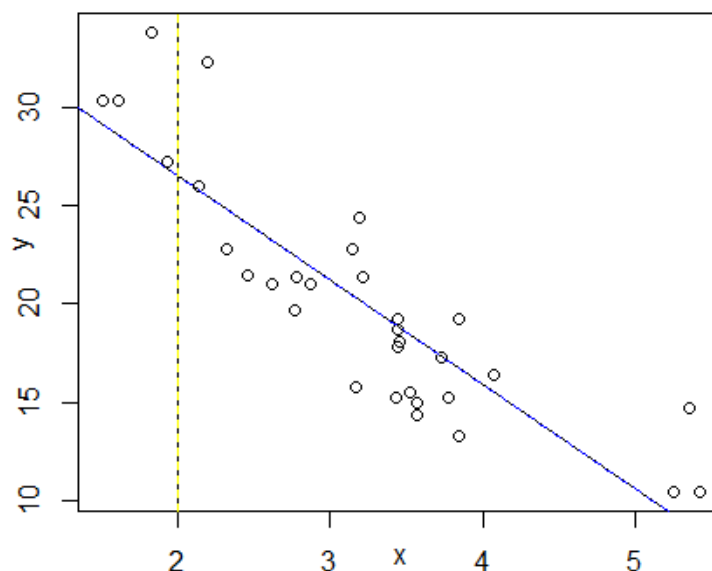
e podemos servir-nos do digrama de dispersão para aí adicionar uma linha usando a função:

```
> abline(valores$coef)
```

A função `abline()` recebe um vetor com dois elementos $c(\beta_0, \beta_1)$ e adiciona uma reta ao gráfico com estes valores de ordenada e inclinação, respetivamente. Se pretendermos apenas uma reta vertical no gráfico devemos indicar

```
> abline(v=<valor>)
```

Figura 2.1: Regressão linear simples



Também podemos calcular o coeficiente de correlação linear de Pearson entre as variáveis.

```
> cor(x, y)
[1] -0.8676594
```

O sinal de dólar, \$, ou colchetes duplos e um número ou nome servem para ter acesso à lista. Assim para a nossa variável de estudo podemos obter o rendimento (a primeira coluna) como uma lista:

```
> dados$mpg; > dados[['mpg']]; > dados[,1]
```

Podemos não usar nenhuma das listas conforme acima, bastando para isso a utilização da função `attach()`, para prender o conjunto de dados no ambiente.

```
> mpg #atualmente não esta visível
```

Error: Object "mpg" not found

```
> attach(mtcars), # função que permite prender os dados
> y=mpg #A lista mpg agora é visível
```

Também podemos ordenar os dados por ordem crescente ou decrescente recorrendo a uma das instruções seguintes:

```
> ord_cresc<-dados[order(dados$mpg),] #ordem crescente
> ord_decresc<-dados[order(-dados$mpg),] #ordem decrescente
> dados_ord <- sort(dados$mpg, decreasing = FALSE or TRUE)
#crescente ou decrescente
```

2.4. Programação no R

Enquanto o R é muito útil como uma ferramenta para análise de dados, muitos utilizadores por vezes têm necessidade de criar as suas próprias funções. Esta é uma das maiores vantagens do R. Além de ser um programa para análises estatísticas, o R é acima de tudo uma linguagem de programação, com a qual podemos programar nossas próprias funções. Aprender a usar o R, mesmo sem saber construir programas básicos tem sempre vantagens, porém, quando se conciliam essas duas competências pode existir grande benefício.

Tecnicamente o R é uma linguagem de expressão com uma sintaxe bastante simples. Exemplos de alguns operadores utilizados em linguagem R:

Operadores relacionais

- igual: ==
- diferente: !=
- maior/menor que: ><
- maior/menor ou igual: >= <=

Operadores lógicos

- e: &
- ou: |
- negação: !

2.4.1. Estruturas de Controlo de Programação

Segundo Torgo (2009), a linguagem R como qualquer outra linguagem de programação, possui várias instruções destinadas a alterar o curso sequencial normal de execução dos programas.

Execução condicional: `if`

O comando "`if`" permite-nos fazer algo baseado em uma condição. Ele assume duas formas:

i. "`if then`"

```
if (condição) {declaração se condição for verdadeira}
```

ii. "`if then else`"

```
if (condição) {declaração se condição for verdadeira }  
else{declaração se condição for Falsa }
```

Para execuções repetidas usam-se os ciclos: **while**, **repeat** e **for**

O ciclo **while**

A sua forma genérica é,

```
while (condição booleana){bloco de instruções}
```

As instruções no bloco dentro do ciclo podem nunca ser executadas, bastando para isso que a condição seja falsa da primeira vez que o R “chega” ao ciclo.

O ciclo **repeat**

A sua forma genérica é,

```
repeat {bloco de instruções}
```

Se não existe qualquer condição lógica a controlar a execução repetida do bloco de instruções, como no caso do ciclo *while*, o R recorre-se à instrução *break* que permite parar a execução de um processo iterativo.

O ciclo **for**

A instrução "for" permite controlar o número de vezes que um ciclo é executado através de uma variável de controlo que vai tomar uma série de valores predefinidos em cada iteração do ciclo. A sua sintaxe genérica é,

```
for (var in sequencia){bloco de instruções}
```

2.4.2. Funções definidas pelo utilizador

Os objetos do tipo *function* são objetos do R que implementam funções, e que podem ser usados em expressões, em instruções e na implementação de outras funções (Girke, Thomas [s.d]).

O nosso *package*, será criado e irá ficar inserido numa função (*function*).

CAPÍTULO 2

A LINGUAGEM DE PROGRAMAÇÃO R

A criação de uma função é uma atribuição com a forma genérica,

```
> nome_da_função<-function(lista de argumentos){corpo da
função}
```

A sintaxe (`function`) é a ordem utilizada, para dizermos ao R, "*estou criando uma função*".

A lista de argumentos é uma lista, separada por vírgulas, que apresenta os argumentos que serão avaliados pela sua função. O corpo da função vem entre chavetas e é a parte onde escrevemos o "algoritmo" que se deseja.

Ao criar uma função vamos querer guardá-la, e para isso basta atribuir-lhe um nome.

Por exemplo, a função que determina a estatística t

```
> t <- function(n) {
+ mu <- 10; sigma<-5;
+ X <- rnorm(n,mu,sigma)
+ (mean(X) - mu)/(sd(X)/sqrt(n)) }
> t(4)
[1] [1] 1.170853
```

Uma função pode ser definida no R numa única linha de comandos: nesse caso as instruções no corpo da função são separadas por “;”.

Por exemplo, a seguinte função devolve o número de ocorrências do valor k no vetor x .

```
> conta<-function(x, k){n<-length(x); c<-0;
+ for (i in 1:n)
+ {if (x[i]==k) c<-c+1};
+ return(c) }
> conta(k=5, x=c(3,-2,5,4,5,-10,11,5,7))
[1] 3
```

CAPÍTULO 3
ANÁLISE CONJUNTA DE REGRESSÕES LINEARES

3. Análise Conjunta de Regressões Lineares

3.1. Nota histórica da ACR

A iniciar este trabalho vamos referir brevemente a história da **Análise Conjunta de Regressões Lineares (ACR)**.

Segundo Oliveira (2007), os primeiros a desenvolver esta técnica foram Mooers (1921) e Yates & Cochran (1938), onde 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. O mesmo autor refere que nos anos sessenta, Finlay & Wilkinson (1963), Eberhart & Russel (1966) e Perkins & Jinks (1968), deram um grande contributo, para a aplicação da técnica, permitindo a sua utilização na comparação do desempenho de cultivares quanto sujeito a um vasto leque de capacidades produtivas dos pares (local, ano).

Na década de 80, Westcott (1986) e Lin et al. (1986) criticaram o facto de, na ACR não se considerarem variáveis ambientais específicas e para o facto de que a escolha da média de todos os cultivares em cada condição como índice ambiental, poder não verificar a independência entre as variáveis especialmente se o número de cultivares for inferior a quinze. Becker & Leon (1988), ignoraram a crítica, considerando-a como pouco importante.

Segundo Oliveira (2007) na década de noventa, Romagosa & Fox (1993) argumentam que a técnica ACR revela fragilidade quando os ambientes testados são restritos em número e em diversidade ou quando os cultivares incluídos na experiência foram já alvo de forte seleção ou quando não são suficientemente representativos para adequação ao modelo.

Entretanto esta técnica tem um bom desempenho, no caso português, ver, por exemplo, Gusmão (1986b), (1990), Gusmão et al. (1991a), (1991b), (1992a), (1992b), (1993a), (1993b), (1995), Ferreira et al. (1990), Maças et al. (1990), Mexia et al. (1991) e Dias (2000).

Em Portugal, os trabalhos inicialmente foram apresentados por Gusmão (1985), (1986a), (1986b) e (1988) onde a ACR era aplicada a redes de ensaios em blocos casualizados, sendo os índices ambientais correspondentes aos diferentes blocos medidos pelos respectivos rendimentos médios. Por conseguinte, para cada bloco era calculada a média das produções ajustando-se, cultivar a cultivar, regressões lineares das produções médias. Esta técnica tinha uma limitação provocada pela sua aplicabilidade apenas ser possível em situações de blocos completos casualizados e também surgem algumas objeções quanto à consideração das produções médias como valores de uma variável controlada. Assim, Mexia et al. (1999), com a introdução dos índices L_2 , alargaram o campo de utilização da ACR ao caso incompleto. Além disso, o uso de índices ambientais L_2 leva a estimadores de máxima verossimilhança e aos testes de razão de verossimilhança.

Mexia et al. (1996) mostraram que o contorno superior definido pelas regressões ajustadas em conjunto é uma poligonal convexa. Também Mexia et al. (1997), mostraram que o mesmo contorno constituído por segmentos das linhas de regressão ajustada, que contém os rendimentos ajustados mais elevados para os índices ambientais é uma poligonal convexa e em Pereira & Mexia (2008) é dito que cada um destes segmentos corresponderá a um intervalo de variação dos índices ambientais em que os cultivares associados terão o máximo rendimento ajustado.

Entre outros autores, Pinto (2006) e Oliveira (2007) mostraram como utilizar a ACR na condução de planos de melhoramento. Aplicaram os seus resultados ao plano português de melhoramento de trigo mole (1986-2000) e de melhoramento da aveia (1984-2004) respetivamente.

A nível internacional, ressaltamos que Digby (1979) apresentou um método numérico iterativo para o ajustamento de um modelo de regressão conjunta a uma tabela incompleta de duas entradas sem repetições, sem no entanto serem estudadas as propriedades estatísticas dos estimadores obtidos. No trabalho de Ng & Grunwald (1997) são também desenvolvidos métodos iterativos para o ajustamento conjunto de regressões.

Ambos os métodos foram comparados em Ng & Williams (2001), sendo que o de Ng & Grunwald (1997) revela francamente vantajoso no que concerne à velocidade de convergência, para além da vantagem de permitir a realização de inferência.

3.2. Aspetos preliminares

Como referenciado acima, tradicionalmente (caso completo) a ACR era aplicada a redes de ensaios em blocos casualizados estando presentes todos os cultivares em todos os blocos, isto é tem-se:

$$p_{i,1} = p_{i,2} = \dots = p_{i,J} = p_i, \quad i = 1, \dots, b.$$

Atualmente utilizam-se frequentemente modelos com blocos incompletos (Caso Geral), salientando os α designs em que os blocos estão agrupados em superblocos estando cada cultivar presente α vezes em cada superbloco isto é as observações serão afetadas por pesos $p_{i,j}$ atribuídos ao j -ésimo cultivar, $j = 1, \dots, J$, no i -ésimo bloco, $i = 1, \dots, b$.

Deste modo o modelo de regressão comum é

$$Y_{i,j} = \alpha_j + \beta_j x_i + \varepsilon_{i,j}, \quad i = 1, 2, \dots, b, \quad j = 1, \dots, J, \quad (3.2.1)$$

com os parâmetros a estimar:

$$\alpha^J = (\alpha_1, \dots, \alpha_J), \quad \beta^J = (\beta_1, \dots, \beta_J), \quad X^b = (x_1, \dots, x_b)$$

O principal problema de tal modelação é como estimar os parâmetros. No nosso estudo o parâmetro que constitui maior problema é a variável não observável: o índice ambiental. Esta variável controlada, mede, para cada par (local, ano), a respetiva capacidade produtiva. Sendo estes valores desconhecidos, terão de ser estimados, através do algoritmo zigzag, de que falaremos adiante.

Variáveis não observáveis são variáveis para as quais não se dispõe à partida de qualquer medida.

3.2.1. Teorema de Scheffé

Seguiremos a apresentação de Mexia (1995) onde se podem encontrar as demonstrações dos resultados que agora se apresentam.

Dado $\underline{Z} \sim \mathcal{N}(\eta, V)$ tem-se

$$U = (\underline{z} - \eta)^t V (\underline{z} - \eta) \sim \sigma^2 \lambda_c^2$$

com $c = \text{car}(V)$ ou seja a característica da matriz V .

Se \underline{Z} for independente de $S \sim \sigma^2 \lambda_g^2$, tem-se

$$F^2 = \frac{g}{c} \frac{U}{S} \sim F(z|c, g, 0)$$

Sendo $f_{p,c,g}$ o p-ésimo quantil de $F(z|c, g, 0)$ teremos

$$pr \left[(\eta - \underline{z})^t V^t (\eta - \underline{z}) \leq c f_{1-q,c,g} \left(\frac{S}{g} \right) \right] = pr(F^* \leq f_{1-q,c,g}) = 1 - q$$

Obtendo assim um elipsóide de confiança de nível $1 - q$ para η .

Agora estabelecemos

Teorema de Scheffé

Sendo $\bigcap_{\underline{d}}$ todos os vectores com o mesmo número de componentes de η tem – se

$$pr \left[\bigcap_{\underline{d}} \left(|\underline{d}^t \eta - \underline{d}^t \underline{Z}| \leq \sqrt{c f_{1-q,c,g} \underline{d}^t V \underline{d} \frac{S}{g}} \right) \right] = 1 - q \quad (3.2.2)$$

Dem: Observa-se que quando $\underline{d} = \underline{0}$ se tem a desigualdade trivial $0 \leq 0$ pelo que nos basta considerar os $\underline{d} \neq \underline{0}$. Cada um destes vectores é ortogonal a um par de planos paralelos tangentes ao elipsóide de confiança de nível $1 - q$ estando o ponto afim de η entre esses planos se e só se, ver Scheffé (1959),

$$|\underline{d}^t \eta - \underline{d}^t \underline{Z}| \leq \sqrt{c f_{1-q,c,g} \underline{d}^t V \underline{d} \frac{S}{g}}$$

Para completar a demonstração basta-nos observar que um ponto está no interior de um elipsóide se e só se estiver entre todos os pares de planos paralelos tangentes ao elipsóide. ■

Observe-se que quando

$$\sqrt{c f_{1-q,r,g} \underline{d}^t V \underline{d} \frac{S}{g}} < |\rho - \underline{d}^t \underline{Z}|$$

a desigualdade correspondente a \underline{d} não se pode verificar com

$$\rho = \underline{d}^t \eta$$

dizendo-se então que $\rho = \underline{d}^t \eta$ é significativamente diferente de ρ ao nível q .

Por outro lado o teorema de Scheffé dá-nos intervalos de confiança simultâneos

$$\underline{d}^t \underline{Z} - \sqrt{cf_{1-q, cg} \underline{d}^t V \underline{d} \frac{S}{g}}; \underline{d}^t \underline{Z} + \sqrt{cf_{1-q, cg} \underline{d}^t V \underline{d} \frac{S}{g}}$$

Para os parâmetros

$$\rho(\underline{d}) = \underline{d}^t \eta$$

tendo-se $\rho(\underline{d}) = \underline{d}^t \eta$ significativamente diferente de $\rho_{(0)}$ ao nível q se e só se o intervalo de confiança para $\rho(\underline{d})$ não contiver $\rho_{(0)}$.

Em particular com

$$\rho(\underline{\delta}_{i,j}) = \underline{\delta}_i - \underline{\delta}_j = \begin{bmatrix} 0 \\ \dots \\ 1 \\ \dots \\ -1 \\ \dots \\ 0 \end{bmatrix}, i = 1 \neq j = -1$$

tem-se que

$$\rho(\underline{\delta}_{i,j}) = \eta_i - \eta_j; i \neq j$$

quando $\rho(\underline{\delta}_{i,j})$ é significativamente diferente de 0 ao nível q ter-se-á, com probabilidade do erro majorada por q ,

$$\eta_i \neq \eta_j$$

Assim podemos, dado os intervalos de confiança serem simultâneo, com a probabilidade de estarmos certos não inferior a $1 - q$ afirmar que diferem todos os pares tais que

$$\sqrt{cf_{1-q, c, g} \underline{\delta}_{i,j}^t V \underline{\delta}_{i,j} \frac{S}{g}} < |\eta_i - \eta_j|$$

com

$$V = \begin{bmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,n} \end{bmatrix}$$

tem-se

$$\underline{\delta}_{i,j}^t V \underline{\delta}_{i,j} = v_{i,i} - 2v_{i,j} + v_{j,j}$$

3.2.2. Ajustamento aos Mínimos Quadrados Estruturados

Na formulação de um modelo utiliza-se a informação disponível para exprimir, sob a forma de equação, as relações entre as variáveis intervenientes: controladas e dependentes. No caso duma única variável dependente, relacionada linearmente com as variáveis explicativas, os valores destas, correspondentes a uma observação da variável dependente, constituem uma linha da matriz do modelo. Neste caso, a matriz do modelo é conhecida, desde que se conheçam os valores das variáveis explicativas, o que nem sempre se verifica. Por exemplo, em caso da ACR, os índices ambientais que medem a produtividade de cada par (local, ano) terão de ser estimados. Os mínimos quadrados estruturados, surgiram como uma técnica para estimar, os valores das variáveis explicativas, mas também os coeficientes das regressões (Pinto, 2006).

Com θ_b parâmetro de estrutura, as matrizes do modelo serão apresentadas por $\mathbf{X}(\theta_b)$, do tipo de $n \times k$ com $b < n$ e $k < n$, e o vetor dos coeficientes por ω_k . Sendo y_n o vetor das observações, admite-se que o vetor médio $E(y_n)$, pertence ao espaço imagem $\Omega^0 = R[\mathbf{X}(\theta_b^0)]$, com θ_b^0 o verdadeiro valor do parâmetro de estrutura. Obtém-se assim uma família parametrizada de modelos. Para estimarmos conjuntamente as componentes dos valores ω_k e θ_b , minimiza-se

$$S(\omega_k | \theta_{1,b}) = \|y_n - \mathbf{X}(\theta_{1,b})\omega_k\|^2 \quad (3.2.3)$$

em ordem às componentes do vetor ω_k . Sendo $\tilde{\omega}_{1,k}$ o mínimo obtido, minimiza-se em seguida

$$S(\theta_b | \tilde{\omega}_{1,k}) = \|y_n - \mathbf{X}(\theta_b)\tilde{\omega}_{1,k}\|^2 \quad (3.2.4)$$

Na secção 3.3 iremos particularizar a aplicação desta abordagem ao estudo da ACR, interessa-nos verificar a possibilidade de se proceder a um ajustamento de escala para vetores de estrutura. Para esse efeito, sendo \aleph a família das matrizes do modelo, pode-se definir em \aleph uma relação de equivalência ζ , escrevendo

$$\mathbf{X}(\dot{\theta}_b) \zeta \mathbf{X}(\ddot{\theta}_b) \quad (3.2.5)$$

quando se tem

$$R[\mathbf{X}(\dot{\theta}_b)] = R[\mathbf{X}(\ddot{\theta}_b)] \quad (3.2.6)$$

Pode admitir-se que os vetores de estrutura identificam as matrizes do modelo, tendo-se $\mathbf{X}(\dot{\theta}_b) = \mathbf{X}(\ddot{\theta}_b)$ se e só se $\dot{\theta}_b = \ddot{\theta}_b$. Sendo Γ o espaço paramétrico dos vetores de estrutura, poremos $\dot{\theta}_b \zeta \ddot{\theta}_b$, quando (3.2.6) se verificar. Interessa-nos o caso em que ζ é uma relação de homotopia, isto é quando $\forall d$

$$(c\theta_b + d1_b) \zeta \theta_b \quad (3.2.7)$$

sempre que $c \neq 0$. A expressão anterior subentende que, dados $\theta_b \in \Gamma$ e $c \neq 0$, $c\theta_b + d1_b \in \Gamma$. Representando por $\triangleleft v_b$ e $\triangleleft v_b$ respetivamente o mínimo e o máximo das componentes de v_b , podemos no caso dos modelos homotópicos, substituir o mínimo $\tilde{\theta}_{1,b}$, obtido pela minimização de $S(\theta_b | \tilde{\omega}_{1,k})$, por $\tilde{\theta}_{1,b}^*$ tal que

$$\begin{cases} \tilde{\theta}_{1,b}^* \zeta \tilde{\theta}_{1,b} \\ \triangleleft \tilde{\theta}_{1,b}^* = \triangleleft \tilde{\theta}_{1,b} \\ \triangleleft \tilde{\theta}_{1,b}^* = \triangleleft \tilde{\theta}_{1,b} \end{cases} \quad (3.2.8)$$

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

$$\text{Min}_{\omega_k} S(\omega_k | \theta_b) = \|y_{\Omega^\perp}\|^2, \quad (3.2.9)$$

virá

$$\text{Min}_{\omega_k} S(\omega_k | \tilde{\theta}_{1,b}^*) = \text{Min}_{\omega_k} S(\omega_k | \tilde{\theta}_{1,b}). \quad (3.2.10)$$

Agora inicia-se um segundo ciclo com $\theta_b = \tilde{\theta}_{1,b}^*$ e repetir as minimizações até que o valor tomado pela função objetivo estabilize.

3.3. Modelos para blocos incompletos

3.3.1. Função objetivo

Segundo Pereira & Mexia (2002) e Oliveira et al. (2012), obter as estimativas para os coeficientes de regressão (α_j, β_j) e os índices ambientais (x_i) , a função objetivo a ser minimizada deve ser

$$S(\alpha^J, \beta^J, X^b) = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \alpha_j - \beta_j x_i)^2 \quad (3.3.11)$$

em que $Y_{i,j}$ é o rendimento do j -ésimo cultivar no i -ésimo bloco se $p_{i,j} = 1$ e qualquer valor se $p_{i,j} = 0$.

Os pesos $p_{i,j} = 1[0]$ quando o j -ésimo cultivar está presente [ausente] no i -ésimo bloco.

A escolha dos valores iniciais para as componentes do vetor X^b pode ser feita atribuindo-se a cada bloco o valor do rendimento médio para o superbloco em que o bloco está integrado. Sendo $x_i^0, i = 1, 2, \dots, b$ os valores iniciais para os índices ambientais, iniciamos com a minimização de

$$S(\alpha^J, \beta^J) = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \alpha_j - \beta_j x_i^0)^2 \quad (3.3.12)$$

em ordem aos vetores $\alpha^J = (\alpha_1, \dots, \alpha_J)$ e $\beta^J = (\beta_1, \dots, \beta_J)$

Agora nos interessa estabelecer uma relação de equivalência entre trios $(\alpha^J, \beta^J, X^b) \in \mathbb{R}_b \times \mathbb{R}_J \times \mathbb{R}_J$, de vetores.

Lema 3.1 *Pondo-se $(\alpha^J, \beta^J, X^b) \tau (\hat{\alpha}^J, \hat{\beta}^J, \hat{X}^b)$, quando existirem $c \neq 0$ e d tais que*

$$\begin{cases} \hat{\alpha}^J = \alpha^J - \frac{d}{c} \beta^J \\ \hat{\beta}^J = \frac{1}{c} \beta^J \\ \hat{X}^b = cX^b + d_1^b \end{cases} \quad (3.3.13)$$

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

Dem: Considerando o caso em que $c = 1$ e $d = 0$ vem, para todo o trio, $(\alpha^J, \beta^J, X^b) \tau (\alpha^J, \beta^J, X^b)$, e conseqüentemente τ goza da propriedade reflexiva.

Admitamos agora que $(\alpha^J, \beta^J, X^b) \tau (\dot{\alpha}^J, \dot{\beta}^J, \dot{X}^b)$, e existindo $c \neq 0$ e d tais que (3.3.13)

se verifica, então sendo $\dot{c} = c^{-1}$ e $\dot{d} = -\frac{d}{c}$ tem-se

$$\begin{cases} \alpha^J = \dot{\alpha}^J - \frac{\dot{d}}{\dot{c}} \dot{\beta}^J \\ \beta^J = \frac{1}{\dot{c}} \dot{\beta}^J \\ X^b = \dot{c} \dot{X}^b + \dot{d}_1^b \end{cases} \quad (3.3.14)$$

concluindo-se que τ goza da propriedade simétrica.

Finalmente, admitamos que $(\alpha_i^J, \beta_i^J, X_i^b) \tau (\alpha_{i+1}^J, \beta_{i+1}^J, X_{i+1}^b)$, existindo pois $c_i \neq 0$ e d_i tais que (3.3.13) se verifica quando $i=1,2$. Agora com $c = c_1 c_2 \neq 0$ e $d = d_1 c_2 + d_2$, as expressões (3.3.13) verificam-se para o par $[(\alpha_1^J, \beta_1^J, X_1^b), (\alpha_3^J, \beta_3^J, X_3^b)]$ o que mostra que τ goza da propriedade transitiva, completando assim a demonstração do lema. ■

Desta relação de equivalência podemos reter que a função objetivo toma valor constante para todos os trios de vetores numa classe de equivalência τ .

Oliveira (2007) apresenta o estudo da função objetivo, com respectivamente o gradiente e a matriz Hessiana de $S(\alpha^J, \beta^J, X^b)$. O mesmo autor faz uma simplificação da matriz Hessiana, concluindo que após algumas iterações do algoritmo zigzag se obtêm um bom ajustamento final, sendo deste modo obtidos os estimadores dos índices ambientais \tilde{x}_i , $i = 1, 2, \dots, b$ e dos coeficientes de regressão ajustados $(\tilde{\alpha}_j, \tilde{\beta}_j)$, $j = 1, 2, \dots, J$. Considera ainda que a boa qualidade do ajustamento permitirá concluir que os valores dos resíduos

$$r_{i,j} = p_{i,j}(Y_{i,j} - \tilde{\alpha}_j - \tilde{\beta}_j \tilde{x}_i) \approx 0; i = 1, \dots, b; j = 1, \dots, J \quad (3.3.15)$$

serão próximos de zero.

Por conseguinte podemos reescrever as derivadas parciais $\frac{\partial^2 S}{\partial x_i \partial \beta_j}$ de forma mais simplificada

$$\frac{\partial^2 S}{\partial x_i \partial \beta_j} = -2p_{i,j}(Y_{i,j} - \tilde{\alpha}_j - \tilde{\beta}_j \tilde{x}_i) + 2p_{i,j} \beta_j x_i \approx 2p_{i,j} \tilde{\beta}_j \tilde{x}_i \quad (3.3.16)$$

A simplificação (3.3.16) é muito útil, pois possibilita-nos aproximar a matriz hessiana de S pela matriz

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

Seguidamente demonstra-se que a matriz A é definida positiva, o que pressupõe um bom comportamento para o método ACR.

Com $u'_{b+2J} = [a'_b, b'_J, c'_J]$, abaixo temos

$$\begin{aligned} u'_{b+2J} A u_{b+2J} &= 2 \sum_{i=1}^b \left(\sum_{j=1}^J p_{i,j} \tilde{\beta}_j^2 \right) a_i^2 + 4 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \tilde{\beta}_j \tilde{x}_i a_i c_j + 4 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \tilde{\beta}_j a_i b_j \\ &+ 2 \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \right) b_j^2 + 2 \left[2 \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \tilde{x}_i \right) b_j c_j + \sum_{j=1}^J \left(\sum_{i=1}^b p_{i,j} \tilde{x}_i^2 \right) c_j^2 \right] \\ &= 2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} [\tilde{\beta}_j^2 a_i^2 + 2 \tilde{\beta}_j a_i (b_j + c_j \tilde{x}_i) + b_j^2 + 2 b_j c_j \tilde{x}_i + c_j^2 \tilde{x}_i^2] \\ &= 2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} [\tilde{\beta}_j a_i + (b_j + c_j \tilde{x}_i)]^2 > 0 \end{aligned} \quad (3.3.17)$$

O que nos permite dizer que numa vizinhança do ótimo, podemos considerar que a matriz Hess (S) ver Oliveira (2007) é definida positiva, o que nos leva a antecipar bom comportamento do algoritmo zigzag.

3.3.2 Algoritmo zigzag

O algoritmo zigzag, ver Mexia et al. (1999), Pereira & Mexia (2010) e Rodrigues et al. (2011) baseia-se num processo iterativo onde as minimizações alternam entre os coeficientes de regressão e os regressores não observáveis que estão relacionados a uma variedade de características que não podem ser diretamente aferidas, o que se torna difícil identificar, sendo que neste caso podemos considerar os índices ambientais como regressores não observáveis.

Conforme vimos na seção 3.3.1 pretendemos minimizar a função objetivo (3.3.11) sendo $X^b = (x_1, \dots, x_b)$ o vetor de índices ambientais, $\alpha^J = (\alpha_1, \dots, \alpha_J)$ e $\beta^J = (\beta_1, \dots, \beta_J)$ os vetores de coeficientes de regressão. Vimos:

$$S(\alpha^J, \beta^J, X^b) = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \alpha_j - \beta_j x_i)^2$$

em que $Y_{i,j}$ é o rendimento do j -ésimo cultivar no i -ésimo bloco se $p_{i,j}$ e b o número de pares (local, ano).

Como b é o número de pares (local, ano), teremos b índices ambientais, x_1, \dots, x_b , que serão os componentes do vetor de estrutura X^b .

Para o caso completo (ou seja, todos os cultivares estão presentes em cada ambiente), o rendimento médio por bloco pode ser um bom valor inicial para os índices ambientais (Gusmão, 1985).

Quando os blocos incompletos são utilizados, temos uma situação muito conveniente quando α -designs são usados.

No caso α -designs com os valores iniciais para os índices ambientais, toma-se, para cada bloco, a média das produções do superbloco em que está integrado. No pior caso, podem ser tomados quaisquer valores iniciais (Pereira, 2004), mas a escolha de valores perto do índice ambiental acelera o cálculo.

Depois de escolher os valores iniciais para os índices ambientais, a função objetivo ajusta-se pelo método usual dos mínimos quadrados os coeficientes das regressões lineares. Em seguida, esses coeficientes são fixados e novos índices ambientais são obtidos. O processo é repetido até a convergência do algoritmo. Depois de cada iteração, os índices ambientais são redimensionados para que o intervalo de índices ambientais seja mantido inalterado.

A minimização de S pode ser levada a cabo de forma iterativa, ver Mexia et al. (2001). Para aplicar esse algoritmo começamos por escolher adequadamente um vetor inicial $X_0^b = (x_1^0, \dots, x_b^0)$, com a minimização de

$$S_1(\alpha^J, \beta^J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (Y_{i,j} - \alpha_j - \beta_j x_i^0)^2 \quad (3.3.18)$$

em ordem aos vetores α^J e β^J

Para realizar a minimização do vetor X^b observamos que, com $\tilde{\alpha}_1^J$ e $\tilde{\beta}_1^J$ os vetores de componentes $\tilde{\alpha}_j^1, \tilde{\beta}_j^1, j = 1, \dots, J$, a função objetivo, dados estes estimadores, pode ser reescrita como

$$S_1(X^b) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (Y_{i,j} - \tilde{\alpha}_j^1 - \tilde{\beta}_j^1 x_i^1)^2 = \sum_{i=1}^b h_i^1(x_i^1) \quad (3.3.19)$$

onde

$$h_i^1(x) = \sum_{j=1}^J p_{i,j} (Y_{i,j} - \tilde{\alpha}_j^1 - \tilde{\beta}_j^1 x)^2; \quad i = 1, \dots, b \quad (3.3.20)$$

Ao minimizarmos cada um dos $h_i^1(x)$, em ordem a x . É fácil demonstrar que os valores mínimos são o

$$\tilde{x}_i^1 = \frac{\sum_{j=1}^J p_{i,j} \tilde{\beta}_j^1 Y_{i,j} - \sum_{j=1}^J p_{i,j} \tilde{\alpha}_j^1 \tilde{\beta}_j^1}{\sum_{j=1}^J p_{i,j} (\tilde{\beta}_j^1)^2}, \quad i = 1, \dots, b \quad (3.3.21)$$

O modelo é homotópico, como referimos em (3.2.7 e 3.2.8), para podermos manter invariantes o mínimo e o máximo dos índices ambientais utilizados nas varias iterações. Esses extremos serão estabelecidos em (3.3.22) mas também é importante ver a relação de equivalência entre trios de vetores abordada em (3.3.13).

Concluídas as minimizações vamos realizar uma estandardização dos índices ambientais de forma a manter invariante o mínimo e o máximo.

Sejam

$$\begin{cases} m_0 = \min[x_1^0, \dots, x_b^0] \\ M_0 = \max[x_1^0, \dots, x_b^0] \end{cases} \quad \begin{cases} m_1 = \min[\tilde{x}_1^1, \dots, \tilde{x}_b^1] \\ M_1 = \max[\tilde{x}_1^1, \dots, \tilde{x}_b^1] \end{cases} \quad (3.3.22)$$

, tomando-se

$$x_i^1 = m_0 + \frac{M_0 - m_0}{M_1 - m_1} (\tilde{x}_i^1 - m_1), \quad i = 1, \dots, b$$

realiza-se uma estandardização já que quando

$$x_i^1 = m_1 [\tilde{x}_i^1 = M_1]$$

se tem

$$x_i^1 = x [x_i^1 = M_0]$$

Pode em seguida repetir-se o algoritmo e a estandardização tomando como vetor inicial de índices ambientais

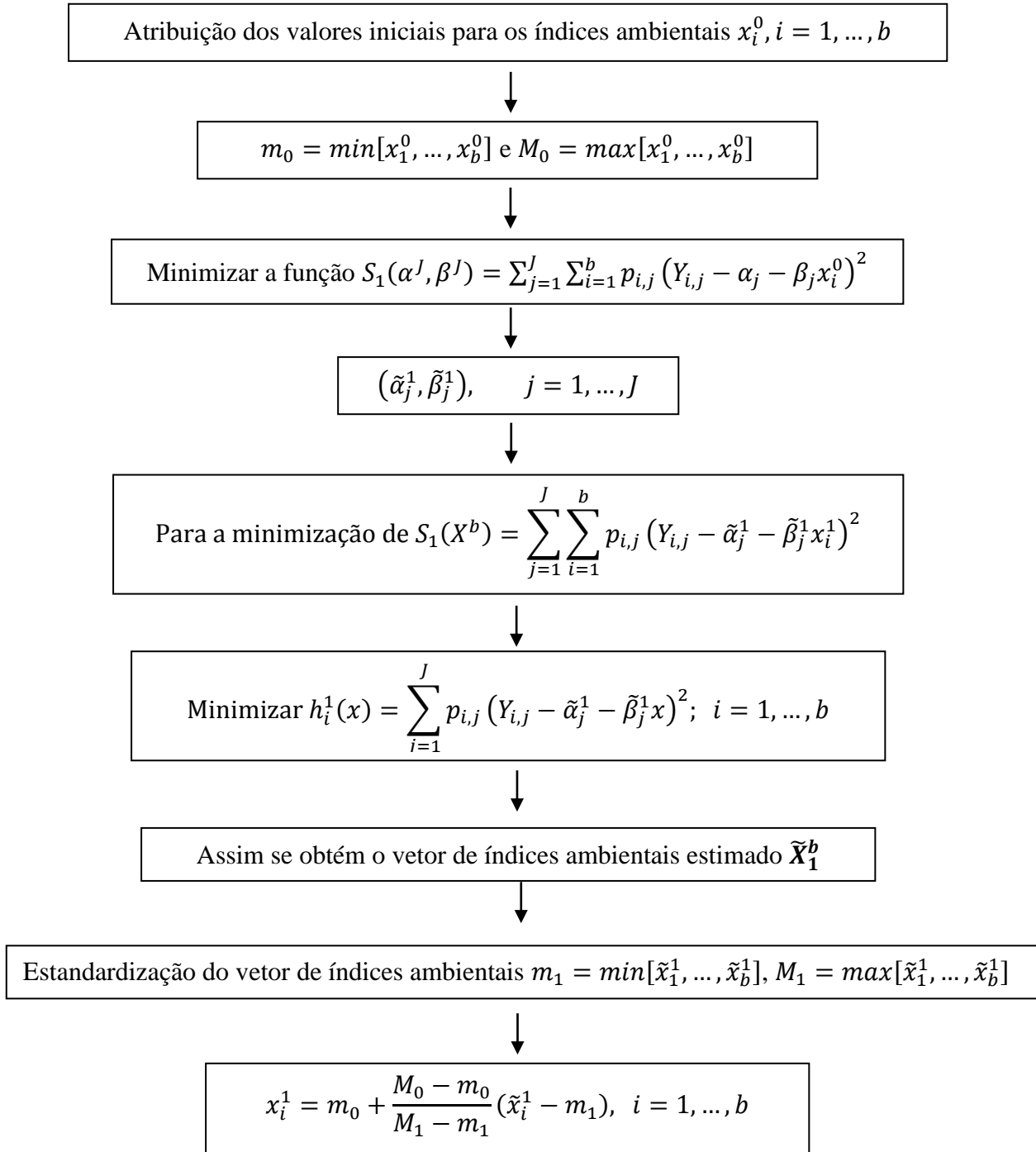
$$X_1^b = \begin{bmatrix} x_1^1 \\ \cdot \\ \cdot \\ \cdot \\ x_b^1 \end{bmatrix}$$

Entra-se assim num processo recursivo que termina quando a soma das somas dos quadrados dos resíduos estabilizar. Na prática o algoritmo tem-se revelado eficiente bastando um número pequeno de ciclos mesmo quando, ver Mexia e Pinto (2003) o algoritmo é aplicado a casos pouco estruturados. Embora não se disponha duma forma teórica de convergência do algoritmo foi possível, realizar a validação estocástica do mesmo.

Supomos que os vetores de rendimento têm componentes normalmente e independentemente distribuídos, de modo que o algoritmo em zigzag levará a estimadores de máxima verossimilhança e nos permitem fazer inferências ao comparar cultivares.

Para realizar o ajuste nós aplicamos o chamado algoritmo de zigzag (ver Pinto (2006) e Pereira (2007)). Este algoritmo é descrito esquematicamente abaixo.

Esquema 1- Algoritmo zigzag



Devemos repetir o procedimento a partir do segundo passo até que as sucessivas somas das somas dos quadrados dos resíduos difiram a menos duma constante suficientemente pequena previamente fixada.

Aplicando sucessivamente a minimização alternativa atrás descrita, obtêm-se uma sequência decrescente de mínimos, contudo só existe um mínimo absoluto se a rede de ensaios for conetada, isto é:

Dados os varios blocos em que as experiências se decompõem, é possível unir todos eles por cadeias de blocos, em que os blocos adjacentes, contêm pelo menos um cultivar em comum.

Pinto (2006) e Pereira (2007), mostram a importância da conexão da rede de ensaios por esta razão, e porque só assim se pode efetuar uma análise integrada da mesma.

Acontece porém que mesmo para as redes de ensaios conetadas, podemos ter mínimos absolutos S que não são equivalentes τ .

Para que os mínimos de S sejam equivalentes τ é necessário e suficiente, garantir que

- i) A rede de ensaios seja conetada;
- ii) $j \leq \text{Int} \left(\frac{n-b}{2} \right) + 1$, com n o número total de parcelas existentes nos b blocos.

Estando a partir daqui nas condições para construir o algoritmo para obtermos o mínimo $S(\tilde{\alpha}^J, \tilde{\beta}^J, \tilde{X}^b)$ da função objetivo $S(\alpha^J, \beta^J, X^b)$.

3.3.3 Análise dos Resíduos

Aqui seguiremos de perto Oliveira (2007) onde apresentaremos uma validação dos ajustamentos usando o algoritmo zigzag, através dum modelo linear para os resíduos.

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 exatos [ajustados] coeficientes de regressão e índices ambientais, teríamos

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

CAPÍTULO 3

ANÁLISE CONJUNTA DE REGRESSÕES LINEARES

onde γ_j , η_j e μ_i são os desvios correspondentes respetivamente 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

$$\overset{\circ}{Y}_{i,j} = Y_{i,j} - (\tilde{\alpha}_j + \tilde{\beta}_j \tilde{x}_i)$$

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

$$\overset{\circ}{Y}_{i,j} = \gamma_j + \tilde{x}_i \eta_j + \tilde{\beta}_j \mu_i + \eta_j \mu_i$$

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

$$\underline{\overset{\circ}{Y}} = \underline{\overset{\circ}{X}} \underline{\overset{\circ}{\theta}} + e$$

onde

$$\begin{cases} \underline{\overset{\circ}{Y}} = [\overset{\circ}{Y}_1 \dots \overset{\circ}{Y}_J] \\ \underline{\overset{\circ}{X}} = [I_J \otimes \overset{\circ}{X} \quad \underline{\tilde{\beta}} \otimes I_b] \\ \underline{\overset{\circ}{\theta}} = (\gamma_1, \eta_1, \dots, \gamma_j, \eta_j, \mu_1, \dots, \mu_b) \end{cases}$$

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

Temos ainda

$$\overset{\circ}{X} = \begin{bmatrix} 1 & \tilde{x}_i \\ \vdots & \vdots \\ 1 & \tilde{x}_b \end{bmatrix}$$

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

$$\underline{\tilde{\theta}} = \left(\overset{\circ}{X} \overset{\circ}{X} \right)^{-1} * \overset{\circ}{X}' \overset{\circ}{Y}$$

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

$$SS_{modelo} = \overset{\circ}{Y}' \overset{\circ}{Y} - \overset{\circ}{Y}' \overset{\circ}{X} \underline{\tilde{\theta}}$$

Para avaliarmos a qualidade do ajustamento obtida pelo algoritmo Zigzag podemos calcular:

$$R^2 = 1 - (SS_{modelo}/SS_{zigzag})$$

onde SS_{zigzag} representa a soma dos quadrados dos resíduos originais obtidos pelo algoritmo zigzag. 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.

3.4. Modelos para blocos completos

Como referenciado na seção 3.2, tradicionalmente, a ACR era aplicada a redes de ensaios em blocos casualizados estando presentes todas os cultivares em todos os blocos, onde o rendimento médio por bloco era um bom valor inicial para a busca dos índices ambientais (Gusmão, 1985).

Neste caso a técnica clássica ajusta cultivar a cultivar regressões lineares das produções nas médias obtidas para cada bloco. Este tipo de procedimento, tem duas desvantagens:

- i) Não ser de todo correto considerar as produções médias como valores duma variável controlada;
- ii) Ser aplicável somente em casos completos.

Verifica-se no entanto que a técnica tem sido aplicada com bons resultados. Podemos usar o algoritmo zigzag nesta técnica, algoritmo descrito acima, ou a técnica de dupla minimização, técnica mais robusta.

Pereira e Mexia (2010) apresentaram um algoritmo alternativo, o algoritmo de dupla minimização, que converge para o mínimo absoluto da função objetivo (3.3.11), afirmando que o mesmo é uma adaptação do algoritmo apresentado pela primeira vez por Fisher e Mackenzie (1923).

3.4.1. Técnica de Dupla Minimização

Quando aplicamos essa técnica obtemos os estimadores condicionais (dados os índices ambientais x_1, \dots, x_b) dos coeficientes de regressão e o mínimo condicional para a soma dos quadrados dos resíduos $\bar{S}(X^b)$. Para escolher os índices ambientais minimizamos $\bar{S}(X^b)$. Ressalta-se que a função objetivo mantém inalterada (temos uma regressão equivalente) se, se fizer uma transformação linear na variável controlada.

Aplicação

Considerando a função objetivo:

$$S(\alpha^J, \beta^J, X^b) = \sum_{i=1}^b \sum_{j=1}^J p_i (Y_{i,j} - \alpha_j - \beta_j x_i)^2$$

onde os pesos p_i correspondem a blocos do mesmo ensaio.

Com o vetor $X_0^b = (x_1^0, \dots, x_b^0)$ de valores iniciais dos índices ambientais, obtém-se

$$\begin{cases} T_x = \sum_{i=1}^b p_i x_i \\ T_{jy} = \sum_{i=1}^b p_i Y_{i,j} \end{cases}, j = 1, \dots, J$$

com as somas

$$\begin{cases} S_{xx} = \sum_{i=1}^b p_i x_i^2 \\ S_{jxy} = \sum_{i=1}^b p_i x_i Y_{i,j} \\ S_{jyy} = \sum_{i=1}^b p_i Y_{i,j}^2 \end{cases}$$

e as somas corrigidas

$$\left\{ \begin{array}{l} s_{xx} = S_{xx} - \frac{T_x^2}{b_j} = \sum_{i=1}^b p_i x_i^2 - \frac{(\sum_{i=1}^b p_i x_i)^2}{\sum_{i=1}^b p_i} = \sum_{i=1}^b p_i \left(x_i - \frac{\sum_{i=1}^b p_i x_i}{\sum_{i=1}^b p_i} \right)^2 \\ s_{jxy} = S_{jxy} - \frac{T_x T_{jy}}{b_j} = \sum_{i=1}^b p_i x_i Y_{i,j} - \frac{\sum_{i=1}^b p_i x_i \cdot \sum_{i=1}^b p_i Y_{i,j}}{\sum_{i=1}^b p_i} \\ \quad = \sum_{i=1}^b p_i \left(x_i - \frac{\sum_{i=1}^b p_i x_i}{\sum_{i=1}^b p_i} \right) \left(Y_{i,j} - \frac{\sum_{i=1}^b p_i Y_{i,j}}{\sum_{i=1}^b p_i} \right); \quad j = 1, \dots, J \\ s_{jyy} = S_{jyy} - \frac{T_{jy}^2}{b_j} = \sum_{i=1}^b p_i Y_{i,j}^2 - \frac{(\sum_{i=1}^b p_i Y_{i,j})^2}{\sum_{i=1}^b p_i} = \sum_{i=1}^b p_i \left(Y_{i,j} - \frac{\sum_{i=1}^b p_i Y_{i,j}}{\sum_{i=1}^b p_i} \right)^2 \end{array} \right.$$

temos agora os estimadores dos coeficientes conhecendo o vetor X^b

$$\begin{aligned} \tilde{\beta}_j(X^b) &= \frac{s_{jxy}}{s_{xx}} = \frac{\sum_{i=1}^b p_i x_i Y_{i,j}}{\sum_{i=1}^b p_i x_i^2}, j = 1, \dots, J \\ \tilde{\alpha}_j(X^b) &= \frac{T_{jy} - \tilde{\beta}_j(X^b) T_x}{b_j} = \frac{\sum_{i=1}^b p_i Y_{i,j} - \tilde{\beta}_j(X^b) \sum_{i=1}^b p_i x_i}{\sum_{i=1}^b p_i} j = 1, \dots, J \end{aligned}$$

As somas dos quadrados dos resíduos são dadas por

$$\begin{aligned} \bar{S}(X^b) &= S(X^b, \tilde{\alpha}^J(X^b), \tilde{\beta}^J(X^b)) = \\ &= \sum_{i=1}^b \sum_{j=1}^J p_i (Y_{i,j} - \tilde{\alpha}_j(x_b) - \tilde{\beta}_j(x_b) x_i)^2 = \sum_{j=1}^J s_{jyy} - \sum_{j=1}^J \frac{s_{jxy}^2}{s_{xx}} \end{aligned}$$

Para Oliveira (2007), na minimização de $\bar{S}(X^b)$, dado que $\sum_{j=1}^J s_{jyy}$ não depende do X^b temos que maximizar

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

Nós podemos nos restringir aos vetores X^b , de tal forma que

$$\left\{ \begin{array}{l} \sum_{i=1}^b p_i x_i = 0 \\ \sum_{i=1}^b p_i x_i^2 = 1 \end{array} \right. \quad (3.4.23)$$

O problema reduz-se à maximização

$$\begin{aligned} M(X^b) &= \sum_{j=1}^J S_{jxy}^2 = \sum_{j=1}^J \left(\sum_{i=1}^b p_i \left(x_i - \frac{\sum_{i=1}^b p_i x_i}{\sum_{i=1}^b p_i} \right) \left(Y_{i,j} - \frac{\sum_{i=1}^b p_i Y_{i,j}}{\sum_{i=1}^b p_i} \right) \right)^2 \\ &= \sum_{j=1}^J \left(\sum_{i=1}^b p_i x_i \left(Y_{i,j} - \frac{\sum_{i=1}^b p_i Y_{i,j}}{\sum_{i=1}^b p_i} \right) \right)^2 \end{aligned}$$

Verificando-se

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

sendo Y_j^{+b} o vetor com 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

$$\begin{aligned} M(X^b) &= \sum_{j=1}^J \left(\sum_{i=1}^b p_i x_i \left(Y_{i,j} - \frac{\sum_{i=1}^b p_i Y_{i,j}}{\sum_{i=1}^b p_i} \right) \right)^2 = \sum_{j=1}^J (x'_b Y_j^{+b})^2 = \sum_{j=1}^J (x'_b Y_j^{+b})(Y_j^{+b} x_b) \\ &= \sum_{j=1}^J x'_b (Y_j^{+b} Y_j'^{+b}) x_b = x'_b \left(\sum_{j=1}^J Y_j^{+b} Y_j'^{+b} \right) x_b \end{aligned}$$

Proposição 3.1. *O vetor X^b que satisfaz as restrições (3.4.23), que maximiza $M(X^b)$ é o primeiro vetor proprio da matriz $(\sum_{j=1}^J Y_j^{+b} Y_j'^{+b})$.*

Dem: Dado que $\sum_{i=1}^b p_i \left(Y_{i,j} - \frac{\sum_{i=1}^b p_i Y_{i,j}}{\sum_{i=1}^b p_i} \right) = 0; j = 1, \dots, J$, então temos

$\left(\frac{\sum_{j=1}^J Y_j^{+b} Y_j'^{+b}}{b^{1/2}} \mathbf{1}_b \right) = \sum_{j=1}^J Y_j^{+b} \left(\frac{Y_j'^{+b}}{b^{1/2}} \mathbf{1}_b \right) = 0$, logo $\frac{1}{b^{1/2}} \mathbf{1}_b$ será um vetor próprio de

$(\sum_{j=1}^J Y_j^{+b} Y_j'^{+b})$ estando-lhe associado um vetor próprio nulo.

Com

$$v' \sum_{j=1}^J Y_j^{+b} Y_j'^{+b} v_b = \sum_{j=1}^J (v'_b Y_j^{+b})(Y_j'^{+b} v_b) = \sum_{j=1}^J (v'_b Y_j^{+b})^2 \geq 0,$$

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

Como $(\sum_{j=1}^J Y_j^{+b})$ não é uma matriz nula, o seu maior valor próprio λ , a que estará associado o primeiro vetor próprio $v_{1,b}$ será positivo (Oliveira, 2007).

Agora $v_{1,b}$ terá de ser ortogonal a $\frac{1}{b^{1/2}} \mathbf{1}_b$, tendo a soma das componentes nula (Giraldes et al., 1995). A demonstração fica completa observando que $\|v_{1,b}\| = 1$, e dado que $v_{1,b}$ satisfaz as restrições (3.4.23) e maximiza $M(X^b)$ na família de todos os vetores, que satisfazem a segunda dessas condições. ■

3.4.2. Testes de razão de verosimilhanças

Na inferência estatística o estimador de máxima verosimilhança é muito importante tendo como função de verosimilhança:

$$L(\alpha^J, \beta^J, X^b, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j \in C_i} p_i (Y_{i,j} - \alpha_j - \beta_j x_i)^2}}{(2\pi)^{N/2} \sigma^N}$$

onde σ^2 é um fator de proporcionalidade das variâncias, $N=bJ$ o número total de observações, C_i conjunto dos cultivares presentes no i -ésimo bloco e os pesos iguais a $p_{i,j}$ com $j = 1, \dots, J$ e $i = 1, \dots, b$.

Admitindo a normalidade e homocedasticidade do vetor dos rendimentos, vem o log-verosimilhança

$$\ln(L(\alpha^J, \beta^J, X^b, \sigma^2)) = -\frac{1}{2\sigma^2} S(\alpha^J, \beta^J, X^b) - \frac{N}{2} (\ln(\sigma^2) + \ln(2\pi))$$

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 o logaritmo dado por:

$$\ln(L(\alpha^J, \beta \mathbf{1}^J, X^b, \sigma^2)) = -\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \alpha_j - \beta x_i)^2 - \frac{N}{2} (\ln(\sigma^2) + \ln(2\pi))$$

deduzir o estimador de σ^2 fazendo

$$\frac{\partial \ln(L(\alpha^J, \beta 1^J, X^b, \sigma^2))}{\partial \sigma^2} = -\frac{1}{2\sigma^4} \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \alpha_j - \beta x_i)^2 - \frac{N}{2\sigma^2}$$

onde $\beta 1^J$ é um vetor de dimensão J com todas as componentes iguais a β .

Igualando a zero e considerando o estimador de máxima verossimilhança para $\hat{\sigma}_k^2$, teremos

$$\hat{\sigma}_k^2 = \frac{1}{N} \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \hat{\alpha}_{j,k} - \hat{\beta}_k \hat{x}_{i,k})^2$$

onde o índice inferior k , indica que são utilizados os estimadores obtidos maximizando a verossimilhança k . Para se obter os estimadores $(\hat{\alpha}_j, \hat{\beta}_j), j = 1, \dots, J$ e $\hat{x}_i, i = 1, \dots, b$ minimizamos

$$S_k(\alpha^J, \beta 1^J, X^b) = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \alpha_{j,k} - \beta_k x_{i,k})^2$$

Aplicando o algoritmo zigzag, com valores iniciais $x_b^1 = \hat{x}_b$, isto é o estimador de máxima verossimilhança sem restrições para os índices ambientais. As primeiras derivadas de S_k em ordem aos parâmetros α_j e β são

$$\frac{\partial S_k}{\partial \alpha_j} = -2 \sum_{i=1}^b p_{i,j} (Y_{i,j} - \alpha_{j,k} - \beta_k x_{i,k}); j = 1, \dots, J$$

$$\frac{\partial S_k}{\partial \beta} = -2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,k} (Y_{i,j} - \alpha_{j,k} - \beta_k x_{i,k}).$$

Igualando a zero

$$\begin{aligned} \frac{\partial S_k}{\partial \alpha_j} = 0 &\Leftrightarrow \sum_{i=1}^b p_{i,j} (Y_{i,j} - \alpha_{j,k} - \beta_k x_{i,k}) = 0 \\ \Leftrightarrow \sum_{i=1}^b p_{i,j} Y_{i,j} - \sum_{i=1}^b p_{i,j} \alpha_{j,k} - \sum_{i=1}^b p_{i,j} \beta_k x_{i,k} &= 0 \end{aligned}$$

então

$$\beta_k = \frac{\sum_{i=1}^b p_{i,j} Y_{i,j} - \sum_{i=1}^b p_{i,j} \alpha_{j,k}}{\sum_{i=1}^b p_{i,j} x_{i,k}}, \text{ com } \alpha_{j,k} = \alpha_1 + d_{j,k}, \quad j = 2, \dots, J$$

e

$$d_{j,k} = \frac{\frac{\sum_{i=1}^b p_{i,j} Y_{i,j}}{\sum_{i=1}^b p_{i,j} x_{i,k}} - \frac{\sum_{i=1}^b p_{i,j} Y_{i,1}}{\sum_{i=1}^b p_{i,j} x_{i,k}}}{\frac{\sum_{i=1}^b p_{i,j}}{\sum_{i=1}^b p_{i,j} x_{i,k}}} = \frac{\sum_{i=1}^b p_{i,j} (Y_{i,j} - Y_{i,1})}{\sum_{i=1}^b p_{i,j}}, \quad j = 2, \dots, J$$

Igualando também a zero as derivadas

$$\begin{aligned} \frac{\partial S_k}{\partial \beta} = 0 &\Leftrightarrow \sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,k} (Y_{i,j} - \alpha_{j,k} - \beta_k x_{i,k}) = 0 \\ &\Leftrightarrow \sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,k} Y_{i,j} - \sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,k} \alpha_{j,k} - \sum_{i=1}^b \sum_{j=1}^J p_{i,j} \beta_k (x_{i,k})^2 = 0 \\ \beta_k &= \frac{\sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,k} Y_{i,j} - \sum_{i=1}^b (\sum_{j=1}^J p_{i,j} x_{i,k}) (\alpha_1 + d_{j,k})}{\sum_{i=1}^b p_{i,j} (x_{i,k})^2} \end{aligned}$$

Igualando os segundos membros das expressões β_k e resolvendo em ordem a α_1 obtém-se,

$$\alpha_1 = \frac{C_1 - C_2}{D_1 - D_2}$$

com

$$\begin{aligned} C_1 &= \frac{\sum_{i=1}^b p_{i,j} \sum_{j=1}^J Y_{i,j} - \sum_{i=1}^b p_{i,j} \sum_{j=1}^J d_{j,k}}{J \sum_{i=1}^b p_{i,j} x_{i,k}} \\ C_2 &= \frac{\sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,k} Y_{i,j} - \sum_{i=1}^b p_{i,j} x_{i,k} \sum_{j=1}^J d_{j,k}}{J \sum_{i=1}^b p_{i,j} x_{i,k}^2} \\ D_1 &= \frac{\sum_{i=1}^b p_{i,j}}{\sum_{i=1}^b p_{i,j} x_{i,k}} \\ D_2 &= \frac{\sum_{i=1}^b p_{i,j} x_{i,k}}{\sum_{i=1}^b p_{i,j} x_{i,k}^2} \end{aligned}$$

Agora são fáceis de obter os valores de $\alpha_2, \dots, \alpha_J$ e de β , correspondentes à primeira iteração passando-se seguidamente à maximização da função $\sum_{i=1}^b h_i(x_i)$ em ordem aos índices ambientais

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

Derivando em ordem a x_i e igualando a zero vem

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

concluindo que

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

Consideremos agora um teste para testar a hipótese

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

Com, α o valor comum das ordenadas na origem, quando H_0 se verifica e k o espaço paramétrico correspondente, teremos agora

$$\hat{\sigma}_k^2 = \frac{1}{bJ} \sum_{i=1}^b p_{i,j} \sum_{j=1}^J (Y_{i,j} - \hat{\alpha}_k - \hat{\beta}_{j,k} \hat{x}_{i,k})^2$$

De novo, com o algoritmo zigzag, podemos minimizar a função

$$S_k(\alpha 1^J, \beta^J, X^b) = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (Y_{i,j} - \alpha_k - \beta_{j,k} x_{i,k})^2$$

onde $\alpha 1^J$ é um vetor de dimensão J com todas as componentes iguais a α .

Para se obter $\hat{\alpha}_k, \hat{\beta}_{j,k}$ e $\hat{x}_{b,k}$

Seja

$$\hat{\beta}_{j,k} = \beta_1 + d_{j,k}, \quad j = 2, \dots, J,$$

com

$$d_{j,k} = \frac{\frac{\sum_{i=1}^b p_{i,j} Y_{i,j} x_{i,k}}{\sum_{i=1}^b p_{i,j} x_{i,k}} - \frac{\sum_{i=1}^b p_{i,j} Y_{i,1} x_{i,k}}{\sum_{i=1}^b p_{i,j} x_{i,k}}}{\frac{\sum_{i=1}^b p_{i,j} x_{i,k}^2}{\sum_{i=1}^b p_{i,j} x_{i,k}}} = \frac{\sum_{i=1}^b p_{i,j} x_{i,k} (Y_{i,j} - Y_{i,1})}{\sum_{i=1}^b p_{i,j} x_{i,k}^2}, \quad j = 2, \dots, J$$

Vindo

$$C_1 = \frac{\sum_{i=1}^b p_{i,j} x_{i,k} \sum_{j=1}^J Y_{i,j} - \sum_{i=1}^b p_{i,j} x_{i,k}^2 \sum_{j=1}^J d_{j,k}}{J \sum_{i=1}^b p_{i,j} x_{i,k}}$$

$$C_2 = \frac{\sum_{i=1}^b p_{i,j} \sum_{j=1}^J Y_{i,j} - \sum_{i=1}^b p_{i,j} x_{i,k} \sum_{j=1}^J d_{j,k}}{J \sum_{i=1}^b p_{i,j}}$$

$$D_1 = \frac{\sum_{i=1}^b p_{i,j} x_{i,k}^2}{\sum_{i=1}^b p_{i,j} x_{i,k}}$$

$$D_2 = \frac{\sum_{i=1}^b p_{i,j} x_{i,k}}{\sum_{i=1}^b p_{i,j} x_{i,k}}$$

Por conseguinte, 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 fim, devemos testar

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

Para

$$\hat{\sigma}_k^2 = \frac{1}{bJ} \sum_{i=1}^b p_{i,j} \sum_{j=1}^J (Y_{i,j} - \hat{\alpha}_k - \hat{\beta}_{j,k} \hat{x}_{i,k})^2$$

queremos minimizar

$$S_k(\alpha^J, \beta^J, X^b) = \sum_{i=1}^b p_{i,j} \sum_{j=1}^J (Y_{i,j} - \alpha - \beta x_{i,k})^2$$

sendo que $a1^J$ tem componente α e $\beta 1^J$ tem componente β

Derivando S_k em ordem aos parametros α e β obtemos

$$\begin{cases} \frac{\partial S_k}{\partial \alpha} = -2 \sum_{i=1}^b p_{i,j} (Y_{i,j} - \alpha - \beta x_{i,k}); j = 1, \dots, J \\ \frac{\partial S_k}{\partial \beta} = -2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,k} (Y_{i,j} - \alpha - \beta x_{i,k}) \end{cases}$$

Igualando as derivadas a zero obtemos um sistema do tipo

$$\begin{cases} \frac{\partial S_k}{\partial \alpha} = 0 \Leftrightarrow a_{11}\alpha + a_{12}\beta = b_1 \\ \frac{\partial S_k}{\partial \beta} = 0 \Leftrightarrow a_{21}\alpha + a_{22}\beta = b_2 \end{cases}$$

cuja solução será

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

em que

$$a_{11} = J \sum_{i=1}^b p_{i,j} \quad ; \quad a_{12} = a_{21} = J \sum_{i=1}^b p_{i,j} x_{i,k} \quad ; \quad a_{22} = J \sum_{i=1}^b p_{i,j} x_{i,k}^2 \quad ;$$

$$b_1 = \sum_{i=1}^b p_{i,j} \sum_{j=1}^J Y_{i,j} \quad e \quad b_2 = \sum_{i=1}^b p_{i,j} x_{i,k} \sum_{j=1}^J Y_{i,j}$$

e

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

sendo que na segunda parte da iteração obtêm-se diretamente

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

3.4.3. Análise de variância

Havendo b locais e n culturas (Tavares, 2013), teremos uma matriz

$$X^b = [x_1, \dots, x_b]$$

em que os vetores coluna contêm os valores dos índices ambientais para as diferentes culturas.

Para o caso completo, tomemos

$$\begin{cases} A_r = 1_r 1_r^t \\ B_r = I_r - \frac{1}{r} A_r \end{cases}$$

sendo $\frac{1}{r} A_r$ e B_r , matrizes simétricas e indempotentes sendo portanto matrizes de projeção ortogonal mutuamente ortogonais já que

$$\frac{A_r}{r} \cdot B_r = B_r \cdot \frac{A_r}{r} = 0_{r \times r}$$

As matrizes de projeção ortogonal mutuamente ortogonais estão associadas a partições ortogonais de subespaços. No nosso modelo teremos

$$\begin{cases} Q_1 = \frac{A_n}{n} \otimes \frac{A_b}{b} \\ Q_2 = B_n \otimes \frac{A_b}{b} \\ Q_3 = \frac{A_n}{n} \otimes B_b \\ Q_4 = B_n \otimes B_b \end{cases}$$

Com espaços imagem $\nabla_1, \nabla_2, \nabla_3$ e ∇_4 mutuamente ortogonais.

O espaço imagem $\Omega^0 = R[X^0]$ de X^0 tem partição ortogonal $\Omega^0 = \bigotimes_{j=1}^4 \nabla_j$

Se os vetores linha de uma matriz M constituem uma base ortonormada para $\nabla = R(B)$ ter-se-á

$$\begin{cases} M^t M = B \\ M^t M = I_h \end{cases}$$

com

$$h = \text{car}(B) = \text{car}(M) = \dim(R(M))$$

No nosso caso temos as matrizes

$$\begin{cases} M_1 = \frac{1}{\sqrt{n}} \mathbf{1}_n^t \otimes \frac{1}{\sqrt{b}} \mathbf{1}_b^t \\ M_2 = Q_n \otimes \frac{1}{\sqrt{b}} \mathbf{1}_b^t \\ M_3 = \frac{1}{\sqrt{b}} \mathbf{1}_b^t \otimes Q_b \\ M_4 = Q_b \otimes Q_b \end{cases} ; \begin{cases} h_1 = 1 \\ h_2 = n - 1 \\ h_3 = b - 1 \\ h_4 = (n - 1)(b - 1) \end{cases}$$

onde, como vimos Q_r é obtida eliminando a primeira linha igual a $\frac{1}{\sqrt{r}} \mathbf{1}_r^t$ numa matriz ortogonal (estandardizada) $p_r^0, r \times r$

Obtidas as matrizes $M_j, j = 1, \dots, 4$ calculam-se os

$$S_j = \|M_j \underline{Z}\|^2, j = 1, \dots, 4$$

Caso se admite ter

$$\underline{Z} \sim \mathfrak{N}(\underline{\lambda}, \sigma^2 \mathbf{1}_{nb})$$

Mexia (1995), $S_j \sim \sigma^2 \lambda_{h_j \delta_j}^2, j = 1, \dots, 4$ com

$$\delta_j = \|M_j \underline{\lambda}\|^2, j = 1, \dots, 4$$

As hipóteses que nos interessa testar relativas a efeitos de locais e de culturas podem ser escritas como:

$$H_{0,j}: M_j \underline{\lambda} = 0, j = 2,3 \text{ ou como } H_{0,j}: \delta_j = 0, j = 2,3$$

Proposição 3.2. *As estatísticas $F_j = \frac{h_4 S_j}{h_j S_4}, j = 2,3$ têm distribuição F com h_j e h_4 graus de liberdade e parâmetros de não centralidade que se anulam quando e só quando $H_{0,j}: \delta_j = 0, j = 2,3$ se verifica.*

Dem: os vetores $M_j \underline{Z}$ e $M_4 \underline{Z}$ tem matriz de covariância cruzada

$$M_j (\sigma^2 I_{nb}) M_4^t = \sigma^2 M_j M_4^t = \mathbf{0}_{k_j \times h_4}, j = 2,3$$

já que como é fácil verificar, as matrizes $M_j M_4^t$, $j = 2,3$ são nulas. Ora os vetores $M_j \underline{Z}$ e $M_4 \underline{Z}$ têm distribuições conjuntas normais, $j = 2,3$, já que

$$\begin{bmatrix} M_j \underline{Z} \\ M_4 \underline{Z} \end{bmatrix} = \begin{bmatrix} M_j \\ M_4 \end{bmatrix} \underline{Z}, \quad j = 2,3$$

tendo matriz de covariância cruzada nula, esses vetores são independentes, logo S_i e S_4 são independentes $j = 2,3$. Como

$$\underline{\lambda} \varepsilon R[X^0] = \bigoplus_{j=1}^t R(M_j^t)$$

tem-se

$M_4 \underline{\lambda} = 0$, logo $\delta_4 = 0$. Assim tem-se $S_j \sim \sigma^2 \lambda_{h_j}^2 \delta_j$, $j = 1,2,3$ independente de $S_4 \sim \sigma^2 \lambda_{h_4}^2$, j donde resulta a parte da tese relativas às distribuições das estatísticas. A parte relativa aos parâmetros da não centralidade resulta diretamente da expressão desses parâmetros.

Observe-se que, com

$$\underline{\psi}_j = M_j \underline{\lambda}, \quad j = 2,3$$

as hipóteses que estamos a considerar podem ser reescritas como

$$H_{0,j}: \underline{\psi}_j = 0, \quad j = 2,3$$

Tratando-se pois de um caso particular de hipóteses consideradas na coluna precedente. O que regulariza o tratamento desta secção é ele assentar na partição ortogonal de R^{nm} nos espaços imagens das matrizes M_j , $j = 1,2,3,4$.

No caso completo a análise pode ser simplificada resolvendo o problema através duma análise de variância para cada um dos fatores.

Sejam

$$\begin{cases} T_{0,i} = \sum_{j=1}^n z_{i,j}, & i = 1, \dots, b \\ T_{0,j} = \sum_{i=1}^b z_{i,j}, & i = 1, \dots, b \\ T_{0,0} = \sum_{i=1}^b \sum_{j=1}^n z_{i,j}, \end{cases}$$

o que nos permite calcular a soma dos quadrados

$$\left\{ \begin{array}{l} S_1^0 = \sum_{i=1}^b \sum_{j=1}^n z^2_{i,j} \\ S_2^0 = \sum_{i=1}^b \frac{T_{0,j}^2}{n} \\ S_3^0 = \sum_{j=1}^n \frac{T_{0,j}^2}{b} \\ S_4^0 = \sum_{j=1}^n \frac{T_{0,0}^2}{nb} \end{array} \right.$$

e a partir daí construir a tabela da ANOVA.

Tabela 3.1: Quadro resumo da ANOVA

Origem da variação	Soma de Quadrados	Graus de liberdade	Quadrado médio	Teste F
Locais	$S_1^0 = S_2^0 - S_4^0$	$(b - 1)$	$Q_1^0 = \frac{S_1^0}{b - 1}$	$F_1 = \frac{S_1^0}{S_E^0}$
Culturas	$S_2^0 = S_3^0 - S_4^0$	$(n - 1)$	$Q_2^0 = \frac{S_2^0}{n - 1}$	$F_2 = \frac{S_2^0}{S_E^0}$
Erro	$S_E^0 = S_1^0 - S_2^0 - S_3^0 - S_4^0$	$(n - 1)(b - 1)$	$Q_E^0 = \frac{S_E^0}{(n - 1)(b - 1)}$	

Se $F_1[F_2]$ é significativo ao nível q e $b > 2 [n > 2]$ deferirão significativamente a esse nível os pares $F_1(i, i')[(j, j')]$ tais que

$$|T_{i,0} - T_{i',0}| > \sqrt{b - 1 f_{1-q, n-1, (n-1)(b-1)} \frac{S_E^0}{(n - 1)(b - 1)}}$$

$$|T_{j,0} - T_{j',0}| > \sqrt{n - 1 f_{1-q, n-1, (n-1)(b-1)} \frac{S_E^0}{(n - 1)(b - 1)}}$$

CAPITULO 4
MÉTODOS DO CONTORNO SUPERIOR

4. Método do Contorno Superior

4.1. Seleção pelo método do Contorno Superior

A utilização do método do contorno superior na comparação e seleção de cultivares foi introduzida por Mexia et al. (1997).

De acordo com Pinto (2006) e Oliveira (2007), a informação fornecida pelo traçado do contorno superior permite facilmente identificar um conjunto de cultivares que se destacam dos restantes para certos valores do índice ambiental.

Consideramos o modelo com j regressões ajustadas

$$\tilde{Y}_j = \tilde{\alpha}_j + \tilde{\beta}_j \theta, \quad j = 1, \dots, J, \quad (4.1.1)$$

uma por cultivar e $\tilde{\theta}^b = [\tilde{\theta}_1, \dots, \tilde{\theta}_b]'$ o vetor dos índices ambientais ajustados e estandardizados tendo $[\theta_{min}, \theta_{max}]$ o intervalo de variação de θ cujos extremos são exatamente o mínimo e o máximo dos índices ambientais ajustados.

Obviamente os cultivares associados as regressões dominantes que são aqueles que correspondem às produções máximas, integrando o contorno superior serão os cultivares dominantes.

A seleção deve dar prioridades aos cultivares que integram o contorno superior e proceder a comparações entre esses cultivares dominantes e todos os restantes, e no caso de algum destes ser significativamente dominado, deverá de imediato ser eliminado do processo de seleção (Pereira e Mexia, 2008).

4.1.1. Geometria do Contorno Superior.

Realizado o ajustamento o contorno superior do conjunto das regressões lineares desempenha um papel muito interessante na seleção dos cultivares. O contorno superior será uma poligonal aberta composta por segmentos de reta.

Vamos atribuir índices às regressões ajustadas, ordenando-as por ordem decrescente dos seus declives ou seja,

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

Assim à regressão com maior declive, será atribuído o índice 1 e à de menor declive o índice J .

Por convenção, verificando-se uma igualdade de declives em duas retas de índices j e j' , $\tilde{\beta}_j = \tilde{\beta}_{j'}$, teremos $\tilde{\alpha}_j > \tilde{\alpha}_{j'}$.

Começamos por observar que dado o par de retas com índices j e j' as mesmas cruzam-se se e só se $\tilde{\beta}_j \neq \tilde{\beta}_{j'}$. A abcissa do ponto de interseção é

$$\theta_{j,j'} = \frac{\tilde{\alpha}_j - \tilde{\alpha}_{j'}}{\tilde{\beta}_j - \tilde{\beta}_{j'}}, j \neq j' \text{ e } j, j' = 1, \dots, J \quad (4.1.2)$$

e as coordenadas desse ponto serão $(\theta_{j,j'}, \tilde{\alpha}_j + \tilde{\beta}_j \theta_{j,j'})$.

E tem-se $\tilde{\alpha}_j + \tilde{\beta}_j \theta \leq [\geq] \tilde{\alpha}_{j'} + \tilde{\beta}_{j'} \theta$ quando $\theta \geq [\leq] \theta_{j,j'}$, isto é, a esquerda [direita] do ponto de interseção tem-se $\tilde{\alpha}_j + \tilde{\beta}_j \theta > \tilde{\alpha}_{j'} + \tilde{\beta}_{j'} \theta$ [$\tilde{\alpha}_j + \tilde{\beta}_j \theta < \tilde{\alpha}_{j'} + \tilde{\beta}_{j'} \theta$]. A partir daqui é fácil estabelecer a

Proposição 4.1. *O contorno superior é uma poligonal convexa.*

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

Para exemplificação da aplicação do método do contorno superior iremos de seguida apresentar exemplos implementados no R.

Nas figuras que se seguem, por conveniência utilizaremos a seguinte simbologia:

$T_M = \theta_{\max}$ = Índice máximo; $T_m = \theta_{\min}$ = Índice mínimo; $T_1 = \bar{\theta}_1$ = Índice 1 e $T_t = \bar{\theta}_t$ = Índice t sendo $t = 2, 3, \dots$

Os exemplos a seguir apresentados tem com base uma função criada no R com a designação `contorno()` descrita no Apêndice A e cuja descrição resumidamente é:

```
> contorno (a, b, x1, x2)
```

onde a é vetor das ordenadas na origem; b é o vetor dos declives; x_1 é o índice ambiental mínimo e x_2 é o índice ambiental máximo.

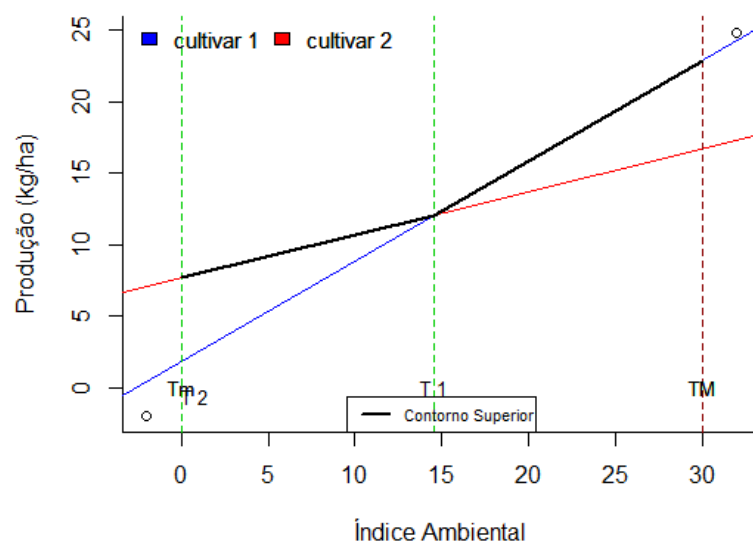
Exemplo 4.1. Interseção de duas retas ajustadas

Como podemos verificar na Figura 4.1, quando dispomos de dois cultivares onde as retas de regressão linear simples têm declives diferentes por exemplo $\tilde{\beta}_j < \tilde{\beta}_{j'}$, e $índice\ j\ [cultivar2] > índice\ j'\ [cultivar\ 1]$, existe um ponto de interseção entre as mesmas.

Sintaxe do Exemplo 4.1

```
> a= c(7.657,1.86) # vetor das ordenadas
> b= c(0.301,0.7) #vetor dos declives
> contorno(a,b,0,30)
O cultivar 1 é dominante entre T1= 14.52882 e TM= 30
O cultivar 2 é dominante entre T 2 = 0 e T 1 = 14.52882
[1] "matriz com declive decrescente"
      a      b
2 1.860 0.700
1 7.657 0.301
```

Figura 4.1: Interseção das retas ajustadas



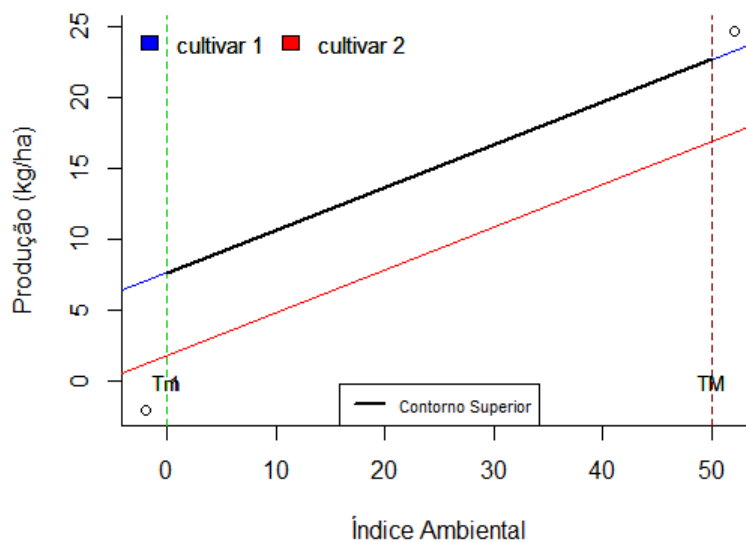
Exemplo 4.2. Retas paralelas

Se não existir nenhum ponto de interseção entre as duas retas com índices j (cultivar 2) e j' (cultivar 1) as mesmas são consideradas paralelas, facto que podemos verificar através da Figura 4.2, onde é dominante aquela com maior ordenada na origem fazendo parte do contorno superior.

Sintaxe do Exemplo 4.2

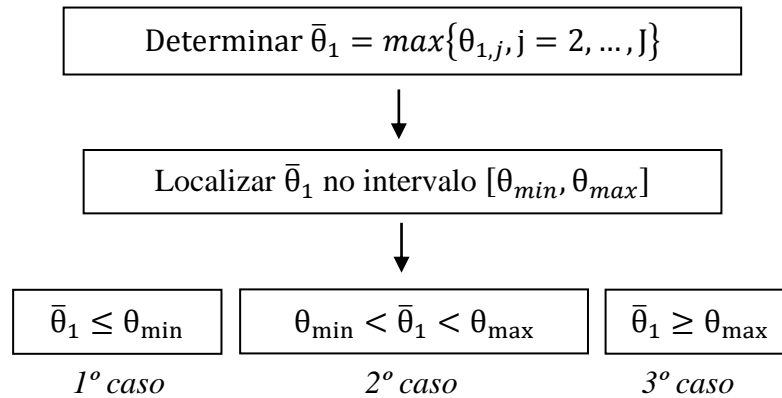
```
> a= c(7.657,1.86) # vetor das ordenadas
> b= c(0.301,0.301) #vetor dos declives
> contorno(a,b,0,50)
O cultivar 1 é dominante entre T1= 0 e TM= 50
Os restantes cultivares não fazem parte do contorno superior!
[1] "matriz com declive decrescente"
      a      b
1 7.657 0.301
2 1.860 0.301
```

Figura 4.2: Retas paralelas



Em Oliveira (2007) apresenta-se a forma para determinação da poligonal que forma o contorno superior, e a discussão dos seguintes três casos mais comuns:

Esquema 2- Determinação do contorno superior



No 1º caso ($\bar{\theta}_1 \leq \theta_{min}$) o contorno superior corresponde exatamente à primeira reta de regressão ajustada $\tilde{\alpha}_1 + \tilde{\beta}_1\theta$, com índice 1 (cultivar 1) e apenas este será dominante conforme se ilustra na Figura 4.3, para $J = 5$.

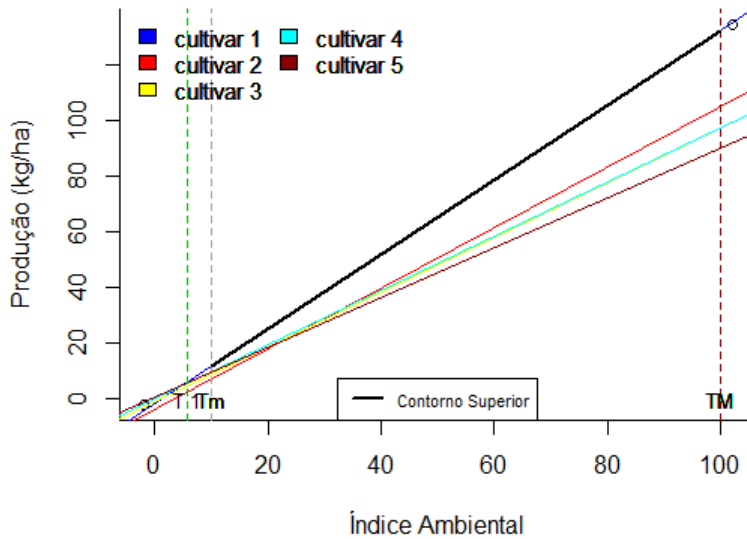
Exemplo 4.3. Contorno superior para $\bar{\theta}_1 \leq \theta_{min}$ com 5 cultivares

Sintaxe do Exemplo 4.3

```

> a= c(-1.74,-4.0,-0.92,0.78,0.06) # vetor das ordenadas
> b= c(1.34,1.09,0.98,0.89,0.97) #vetor dos declives
> contorno(a,b,10,100)
[1] "O cultivar 1 é dominante"
Os restantes cultivares não fazem parte do contorno superior!
[1] "matriz com declive decrescente"
  a    b
1 -1.74 1.34
2 -4.00 1.09
3 -0.92 0.98
5  0.06 0.97
4  0.78 0.89
  
```

Figura 4.3: Traçado do contorno superior quando $\bar{\theta}_1 \leq \theta_{min}$



$$Tm = \theta_{min} = 10 \text{ e } TM = \theta_{max} = 100$$

Exemplo 4.4. Contorno superior para $\theta_{min} < \bar{\theta}_1 < \theta_{max}$ com 5 cultivares

No 2º caso ($\theta_{min} < \bar{\theta}_1 < \theta_{max}$) o contorno superior deverá ser visto nos intervalos :

No intervalo $[\bar{\theta}_1, \theta_{max}]$ com $\tilde{\alpha}_j + \tilde{\beta}_j\theta < \tilde{\alpha}_1 + \tilde{\beta}_1\theta$, $j = 2, \dots, J$ tem-se que o cultivar 1 é dominante. Seja j o índice do cultivar cuja reta de regressão interseta a reta do cultivar com índice 1 no ponto de abscissa $\bar{\theta}_1$. No intervalo $[\theta_{min}, \bar{\theta}_1]$ o cultivar com índice j é dominante. Neste mesmo intervalo haverá para além do cultivar j mais cultivares dominantes e para se averiguar da existência de mais cultivares dominantes terá de se obter

$$\bar{\theta}_2 = \max\{\theta_{j,j'}, j' = j + 1, \dots, J\} \tag{4.1.3}$$

Na Figura 4.4 é visível que no intervalo $[\bar{\theta}_2, \bar{\theta}_1]$ com $[\tilde{\alpha}_j + \tilde{\beta}_j\theta] \tilde{\alpha}_5 + \tilde{\beta}_5\theta < [\tilde{\alpha}_j + \tilde{\beta}_j\theta] \tilde{\alpha}_3 + \tilde{\beta}_3\theta$ tem-se que o cultivar 3 [índice j'] é dominante. Para $\theta_{min}, \bar{\theta}_2$ tem-se $\tilde{\alpha}_j + \tilde{\beta}_j\theta \geq \tilde{\alpha}_j + \tilde{\beta}_j\theta$, o cultivar 5 [índice j] é dominante. Neste seguimento para o intervalo $[\theta_{min}, \bar{\theta}_2]$ terá de se proceder do mesmo modo para determinar $\bar{\theta}_3$ (caso haja) e assim sucessivamente até que se tenha o contorno superior totalmente definido entre $[\theta_{min}, \theta_{max}]$.

Sintaxe do Exemplo 4.4

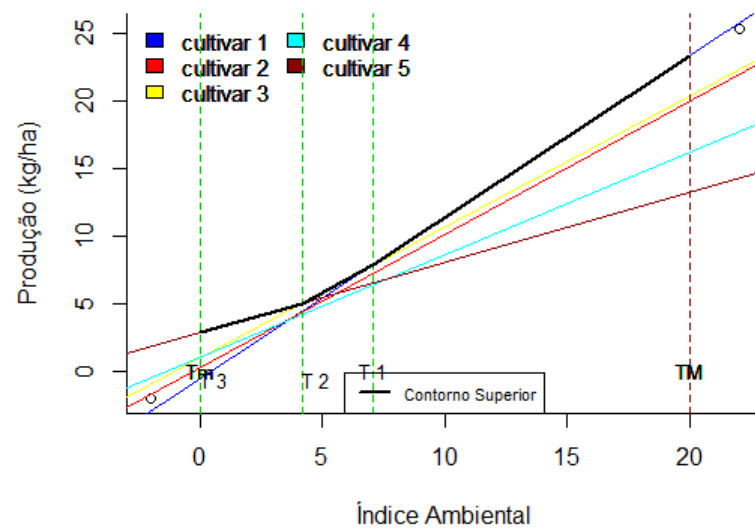
```

> a= c(-0.64,0.22,0.98,2.87,0.99) # vetor das ordenadas
> b= c(1.20,0.99,0.76,0.52,0.97) #vetor dos declives
> contorno(a,b,0,20)
O cultivar 1 é dominante entre T1= 7.086957 e TM= 20
O cultivar 3 é dominante entre T2 = 4.177778 e T1 = 7.086957
O cultivar 5 é dominante entre T3 = 0 e T2 = 4.177778
Os restantes cultivares não fazem parte do contorno superior!

[1] "matriz com declive decrescente"
      a      b
1 -0.64 1.20
2  0.22 0.99
5  0.99 0.97
3  0.98 0.76
4  2.87 0.52

```

Figura 4.4: Traçado do contorno superior para $\theta_{min} < \bar{\theta}_1 < \theta_{max}$



$$T_m = \theta_{min} = 0 \text{ e } T_M = \theta_{max} = 20$$

Exemplo 4.5. Contorno superior para $\bar{\theta}_1 \geq \theta_{max}$ com 4 cultivares

No 3º caso ($\bar{\theta}_1 \geq \theta_{max}$) a reta de regressão com índice 1 não fará parte do contorno superior e portanto o cultivar associado nunca será dominante conforme se ilustra na Figura 4.5, com 4 cultivares.

Por conseguinte temos que trabalhar com o conjunto de retas com índices $j' = 2, \dots, J$ e determinar

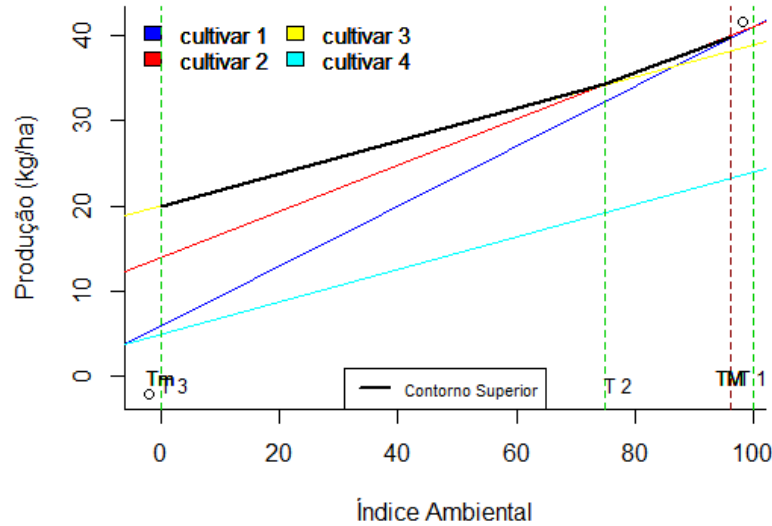
$$\bar{\theta}_2 = \max\{\theta_{2j}, j = 3, \dots, J\} \quad (4.1.4)$$

Caso

- $\bar{\theta}_2 \in [\theta_{min}, \theta_{max}]$, estar-se-á numa situação idêntica à definida pelo 2º caso;
- $\bar{\theta}_2 > \theta_{max}$ o cultivar com índice 2 não será dominante passando a trabalhar-se apenas com um conjunto de retas de regressão com índices $j' = 3, \dots, J$ e assim sucessivamente até estarmos na situação definida no 1º ou no 2º caso.

Sintaxe do Exemplo 4.5

```
> a= c(20,5,6,14) # vetor das ordenadas
> b= c(0.19,0.19,0.35,0.27) # vetor dos declives
> contorno(a,b,0,96)
[1] "O cultivar 1 não faz parte do contorno superior"
   O cultivar 2 é dominante entre T 2 = 75 e T 1 = 96
   O cultivar 3 é dominante entre T 3 = 0 e T 2 = 75
Os restantes cultivares não fazem parte do contorno superior!
[1] "matriz com declive decrescente"
   a    b
3  6  0.35
4 14  0.27
1 20  0.19
2  5  0.19
```

Figura 4.5: Traçado do contorno superior para $\bar{\theta}_1 \geq \theta_{max}$ 

$$T_m = \theta_{min} = 0 \text{ e } T_M = \theta_{max} = 96$$

No 2º caso podem ocorrer diversas possibilidades, conforme passaremos a descrever nas proposições seguintes:

Proposição 4.2. Se $\bar{\theta}_1 = \theta_{1,j'}$, e para $j \neq 1, j', \theta_{1,j'} > \theta_{1,j}$ o cultivar j' é dominante.

Dem: Tendo-se $j \neq 1, j'$ e para $\theta_{1,j} < \theta < \theta_{1,j'}$ verificam-se as desigualdades $\tilde{\alpha}_j + \tilde{\beta}_j \theta > \tilde{\alpha}_1 + \tilde{\beta}_1 \theta > \tilde{\alpha}_j + \tilde{\beta}_j \theta$ e portanto o cultivar com índice j' é dominante. ■

Suponhamos que em $\theta = \bar{\theta}_1$ a reta de índice 1 é interseçada por mais do que uma reta de regressão. Seja C o conjunto dos índices dessas retas e j' o maior índice do conjunto C.

Estabeleçamos a

Proposição 4.3. Se j' é o maior índice para o qual $\bar{\theta}_1 = \theta_{1,j'}$, o cultivar com índice j' é dominante.

Dem: Esta estabelece que, das retas que interseçam a primeira ($\tilde{\alpha}_1 + \tilde{\beta}_1 \theta$), em $\bar{\theta}_1$ irá pertencer o contorno superior à esquerda de $\bar{\theta}_1$ aquela que tiver menor inclinação. Podemos ver que, devido à interseção das retas com índices 1 e j' , temos $\tilde{\beta}_{j'} < \tilde{\beta}_1$. Além disso se $1 < j < j'$, então $\tilde{\beta}_{j'} < \tilde{\beta}_j$, $j = 2, \dots, j' - 1$. Se $j \in C - \{j'\}$, para $\theta > \bar{\theta}_1$, vem

$\tilde{\alpha}_1 + \tilde{\beta}_1\theta > \tilde{\alpha}_j + \tilde{\beta}_j\theta > \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta$, verificando-se a desigualdade contrária se $\theta < \bar{\theta}_1$. Caso $j' \in C$, temos $\theta_{1,j} < \theta_{1,j'} = \bar{\theta}_1$ e para qualquer $\theta \in]\theta_{1,j}, \theta_{1,j'}[$ verifica-se a desigualdade $\tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta > \tilde{\alpha}_1 + \tilde{\beta}_1\theta > \tilde{\alpha}_j + \tilde{\beta}_j\theta$. Portanto a esquerda de $\bar{\theta}_1$, constata-se que existe um intervalo de dominância, para o cultivar com índice j' . ■

Proposição 4.4. *O cultivar com índice j' não é dominante, quando $1 < j < j'$ com j' o maior índice no conjunto C .*

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

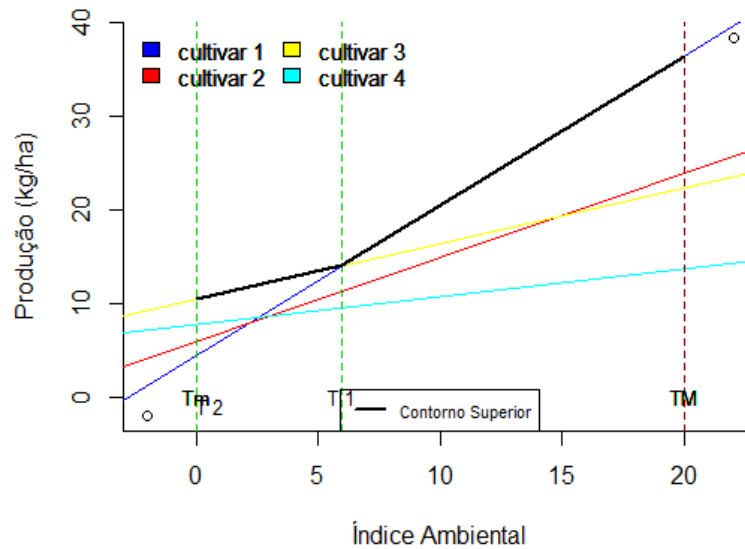
Exemplo 4.6. Contorno superior com dois cultivares dominantes

A Figura 4.6 mostra dois cultivares dominantes com $j=4$ exemplificando a situação prevista pela proposição 4.4.

Sintaxe do Exemplo 4.6

```
> a= c(7.657,5.86,4.4,10.4) # vetor das ordenadas
> b= c(0.301,0.9,1.6,0.6) #vetor dos declives
> contorno(a,b,0,20)
O cultivar 1 é dominante entre T1= 6 e TM= 20
O cultivar 3 é dominante entre T2 = 0 e T1 = 6
Os restantes cultivares não fazem parte do contorno superior!
[1] "matriz com declive decrescente"
      a      b
3  4.400 1.600
2  5.860 0.900
4 10.400 0.600
1  7.657 0.301
```

Figura 4.6: Contorno superior com dois cultivares dominantes



$$T_m = \theta_{min} = 0 \text{ e } T_M = \theta_{max} = 20$$

4.1.2. Métodos de comparação de cultivares

Como afirmamos atrás, os cultivares do contorno superior são chamados dominantes e devem ser selecionados. Os restantes devem ser comparados com os dominantes, a fim de confirmar se eles são significativamente "dominados" em todo o intervalo de dominância dos índices ambientais $[\theta_{min}, \theta_{max}]$.

Pereira e Mexia (2003a, 2008), fizeram uma abordagem a este assunto admitindo a utilização de índices ambientais L_2 ajustados.

Seja X^b o vetor de índices ambientais L_2 ajustado, e x_j uma submatriz de $X = [1^b: X^b]$ cujas linhas correspondem aos blocos que contêm o cultivar j e D_j a matriz diagonal com os pesos, p_i , para os blocos, $j, j = 1, \dots, J$. Então tomamos

$$\begin{bmatrix} \tilde{\alpha}_j \\ \tilde{\beta}_j \end{bmatrix} = (X^T D_j X)^{-1} X^T D_j Y_j^b, \quad j = 1, \dots, J \quad (4.1.5)$$

com Y_j^b o vetor de rendimento para o j -ésimo cultivar.

Assumindo que

$$\begin{bmatrix} \tilde{\alpha}_j \\ \tilde{\beta}_j \end{bmatrix} \sim N \left(\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix}; \sigma^2 W_j \right), \quad j = 1, \dots, J \quad (4.1.6)$$

Independente de $S \sim \sigma^2 \chi_g^2$, onde $W_j = (X_j^T D_j X_j)^{-1}$, $j = 1, \dots, J$, e $g = \sum_{j=1}^J b_j - 2J$ com b_j número de blocos em que está presente o j -ésimo cultivar.

Vamos começar por um caso geral apresentando um teste de hipóteses que se destina a comparar para um determinado valor θ_0 do índice ambiental, uma ou mais das funções lineares $U_j(1, \theta)$, com uma delas, $U_{j'}(1, \theta)$, tomando como padrão ver Pinto (2006).

Com $\theta = \theta_0$, teremos para o par (j, j') as hipóteses testadas e alternativas.

$$\begin{cases} H_{0,j,j'}(\theta_0): U_j(\theta_0) = U_{j'}(\theta_0) \\ H_{1,j,j'}(\theta_0): U_j(\theta_0) < U_{j'}(\theta_0) \end{cases} \quad (4.1.7)$$

com

$$U_j(\theta_0) = \alpha_j + \beta_j \theta_0, \quad j, j' = 1, \dots, J \text{ e } j \neq j'$$

Assim, para testar as hipóteses consideradas deverá ser usada a estatística de teste “t”

$$t_{j,j'}(\theta_0) = \frac{\tilde{U}_j(\theta_0) - \tilde{U}_{j'}(\theta_0)}{\sqrt{\frac{S}{g} [k_j(\theta_0) + k_{j'}(\theta_0)]}} \quad (4.1.8)$$

com

$$\begin{cases} k_l(\theta_0) = [1 \ \theta_0] W_l \begin{bmatrix} 1 \\ \theta_0 \end{bmatrix}; \quad l = 1, \dots, J \\ \tilde{U}_l(\theta_0) = \tilde{\alpha}_l + \tilde{\beta}_l \theta_0 \end{cases} \quad (4.1.9)$$

Uma vez que $\tilde{U}_l(\theta_0)$ é normal com valor de média $U_l(\theta_0)$ e variância $k_l(\theta_0)$, $l = 1, \dots, J$, independente entre si e de $S \sim \sigma^2 \chi_g^2$ vimos que, quando $H_{0,j,j'}(\theta_0)$ se verifica $t_{j,j'}(\theta_0)$ tem distribuição t central com g graus de liberdade.

De forma particular esta comparação deve ser feita sobre os extremos no intervalo de dominância dos cultivares (Pereira e Mexia, 2003b).

Sendo j' o cultivar dominante no intervalo $[\theta^{j'_{min}}, \theta^{j'_{max}}]$, teremos em qualquer dos casos $\beta_{j'} > \beta_j$ a diferença $(\tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta) - (\tilde{\alpha}_j + \tilde{\beta}_j\theta)$ é mínima no intervalo $\theta^{j'_{min}} \leq \theta \leq \theta^{j'_{max}}$ para $\theta = \theta^{j'_{min}}$.

$\beta_{j'} < \beta_j$ a diferença $(\tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta) - (\tilde{\alpha}_j + \tilde{\beta}_j\theta)$ é mínima no intervalo $\theta^{j'_{min}} \leq \theta \leq \theta^{j'_{max}}$ para $\theta = \theta^{j'_{max}}$.

Assim, para comparar os cultivares j, j' , por exemplo para

$j' < j$, se $\beta_{j'} > \beta_j$, significando que no intervalo as retas estão mais próximas quando $\theta = \theta^{j'_{min}}$, somos levados a testar

$$H_{0,j,j'}: \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta^{j'_{min}} = \tilde{\alpha}_j + \tilde{\beta}_j\theta^{j'_{min}}$$

contra

(4.1.10)

$$H_{1,j,j'}: \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta^{j'_{min}} > \tilde{\alpha}_j + \tilde{\beta}_j\theta^{j'_{min}}$$

Para $j' > j$, se $\beta_{j'} < \beta_j$, significando que no intervalo as retas estão mais próximas quando $\theta = \theta^{j'_{max}}$, somos levados a testar

$$H_{0,j,j'}: \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta^{j'_{max}} = \tilde{\alpha}_j + \tilde{\beta}_j\theta^{j'_{max}}$$

contra

(4.1.11)

$$H_{1,j,j'}: \tilde{\alpha}_{j'} + \tilde{\beta}_{j'}\theta^{j'_{max}} > \tilde{\alpha}_j + \tilde{\beta}_j\theta^{j'_{max}}$$

Se o primeiro [último] cultivar é dominante, não haverá cultivares com maior [menor] inclinação e as únicas hipóteses a serem testadas são as descritas anteriormente.

Se pretendermos alcançar um maior nível de robustez para comparar as $U_j(\theta^{j'_{max}}) [U_j(\theta^{j'_{min}})]$, $j = 1, \dots, J$ podemos ainda utilizar os vários métodos de comparação múltipla, como de Scheffé, de Bonferroni e de Tukey que passamos a descrever.

4.1.2.1. Comparação múltipla de Scheffé

Ao usar o método de Scheffé (1959, págs. 66:70), representando por $F_{1-\alpha, r, g}$ o valor crítico da distribuição F de Snedecor central ao nível de significância α , com r e g graus de liberdade, $U_1(\theta_{max}^{j'}), \dots, U_{j'-1}(\theta_{max}^{j'})$, de tal forma que

$$\frac{|\tilde{U}_j(\theta_{max}^{j'}) - \tilde{U}_{j'}(\theta_{max}^{j'})|}{\sqrt{\frac{S}{g} [k_j(\theta_{max}^{j'}) + k_{j'}(\theta_{max}^{j'})] \cdot (j' - 1) \cdot F_{1-\alpha, j'-1, g}}} > 1, \quad j = 1, \dots, j' - 1$$

$\tilde{U}_{j'}(\theta_{max}^{j'})$ será, ao nível α , conjunta e significativamente superiores aos $\tilde{U}_j(\theta_{max}^{j'})$ tais que $j < j'$.

Por outro lado, $U_{j'+1}(\theta_{min}^{j'}), \dots, U_J(\theta_{min}^{j'})$ as quais

$$\frac{|\tilde{U}_j(\theta_{min}^{j'}) - \tilde{U}_{j'}(\theta_{min}^{j'})|}{\sqrt{\frac{S}{g} [k_j(\theta_{min}^{j'}) + k_{j'}(\theta_{min}^{j'})] \cdot (J - j') \cdot F_{1-\alpha, J-j', g}}} > 1, \quad j = j' + 1, \dots, J$$

$\tilde{U}_{j'}(\theta_{min}^{j'})$ será, ao nível α , conjunta e significativamente superiores aos $\tilde{U}_j(\theta_{min}^{j'})$ tais que $j > j'$.

4.1.2.2. Método de comparação múltipla de Bonferroni

Quando estamos interessados em inferir sobre um número fixo de contrastes e em casos em que há possibilidade de melhorar os resultados obtidos pelo método de comparação de múltipla de Scheffé, podemos usar a desigualdade de Bonferroni.

Aqui o objetivo é comparar pares de valores, ver Seber (2003, págs. 120:125) podemos substituir nas expressões anteriores $\sqrt{(j' - 1) \cdot F_{1-\alpha, j'-1, g}}$ e $\sqrt{(J - j') \cdot F_{1-\alpha, J-j', g}}$ pelos quantis de distribuição t de student com g graus de liberdade para as probabilidades $1 - \frac{\alpha}{2 \cdot (j'-1)}$ e $1 - \frac{\alpha}{2 \cdot (J-j')}$ respetivamente.

Quando usamos o método de comparação múltipla de Bonferroni $U_1(\theta_{max}^{j'}), \dots, U_{j'-1}(\theta_{max}^{j'})$, tal que

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

$\tilde{U}_{j'}(\theta_{max}^{j'})$ será, ao nível $\frac{\alpha}{2.(j'-1)}$, conjunta e significativamente superiores aos $\tilde{U}_j(\theta_{max}^{j'})$ tais que $j < j'$.

Por outro lado, $U_{j'+1}(\theta_{min}^{j'}), \dots, U_J(\theta_{min}^{j'})$ as quais

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

$\tilde{U}_j(\theta_{min}^{j'})$ será, ao nível $\frac{\alpha}{2.(J-j')}$, conjunta e significativamente superiores aos $\tilde{U}_{j'}(\theta_{min}^{j'})$ tais que $j > j'$.

4.1.2.2. Método de comparação múltipla de Tukey

Os métodos de comparação múltipla (Scheffé e Bonferroni) podem ser utilizados tanto no caso completo bem como no caso incompleto. No entanto, para o caso completo, pode também ser utilizado o método de Tukey ver Scheffé (1959, pág. 73).

Quando usamos o método de comparação múltipla de Tukey, $U_1(\theta_{max}^{j'}), \dots, U_{j'-1}(\theta_{max}^{j'})$, tal que

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

Onde $T_{1-\alpha, j', g}$ é o quantil da distribuição t de student com g graus de liberdade ao nível α de significância.

$\tilde{U}_{j'}(\theta_{max}^{j'})$ será, ao nível α , conjunta e significativamente superiores aos $\tilde{U}_j(\theta_{max}^{j'})$ tais que $j < j'$.

Por fim, $U_{j'+1}(\theta_{min}^{j'}), \dots, U_J(\theta_{min}^{j'})$ para os quais

$$\frac{|\tilde{U}_j(\theta_{min}^{j'}) - \tilde{U}_{j'}(\theta_{min}^{j'})|}{T_{1-\alpha, J-j'+1, g} \sqrt{\frac{s}{g} k(\theta_{min}^{j'})}} > 1, \quad j = j' + 1, \dots, J$$

$\tilde{U}_{j'}(\theta_{min}^{j'})$ será, ao nível α , conjunta e significativamente superiores aos $\tilde{U}_j(\theta_{min}^{j'})$ tais que $j > j'$.

Ressalta-se que no caso completo $k_l(\theta_0) = k(\theta_0)$ uma vez que as matrizes $X_l, l = 1, \dots, J$ são todos iguais a $X = [1^b: X^b]$.

Para medir a eficácia da seleção, podemos usar os rácios

$$\begin{cases} r_1 = \frac{\text{Número de cultivares dominantes}}{\text{Número de cultivares}} \\ r_2 = \frac{\text{Número de cultivares não dominadas}}{\text{Número de cultivares}} \end{cases}$$

A fim de se obter a proporção de cultivares dominantes e não dominados, respetivamente.

A potência ACR na seleção de cultivares aumenta com a diminuição de r_1 .

Outras técnicas, como o controle da taxa de falsa descoberta pode ser usada para estas comparações (Pereira e Mexia, 2008).

**CAPITULO 5 - PACKAGE EM LINGUAGEM R PARA A
EXECUÇÃO AUTOMÁTICA DA TÉCNICA ACR.**

5. Package em linguagem R para a execução automática da técnica ACR.

O *package* ‘jra’ é formado por quatro funções principais, pela ordem, **zigzag**(), **R2**()’ **Contorno**() e **compMult**().

Foi desenvolvido para responder aos objetivos da dissertação após uma revisão da ACR e tendo em vista uma possível posterior utilização por parte de investigadores e profissionais interessados nesta área de estudo.

De seguida será apresentada uma descrição relativa à documentação e estrutura do projeto de elaboração de um *package*:

5.1. Documentação do package

Package, ‘jra’

Title: Joint Regression Analyses

Version: 0.0.1

Date: a definir

Description: Aplica técnicas usadas em Análise Conjunta de Regressões para comparação e seleção de cultivares ensaiados.

Imports: MASS, graphics

Depends: R (>= 3.2.1)

License:

LazyData: true

NeedsCompilation: no

Author: Adriano Arcanjo Monteiro, adri_arcanjo@hotmail.com

*zigzag**Algoritmo zigzag*

Description

Realiza o algoritmo zigzag para minimizar iterativamente a função objetivo do modelo

$E(Y_{ij}) = \alpha_j + \beta_j x_i$ onde x corresponde a um vetor de variáveis não observáveis. Veja o capítulo 3 para maiores detalhes.

Usage

```
zigzag(y, nomes = NULL, precision = 1e-06)
\method{residuals}{zigzag}(object, ...)
\method{print}{zigzag}(x, ...)
```

Arguments

<code>y</code>	matriz de cultivares. Deve conter uma coluna para cada cultivar e uma linha para cada bloco. Deve ser da classe <code>data.frame</code> ou <code>matrix</code> .
<code>nomes</code>	vetor com nomes dos cultivares. Deve ser fornecido caso o objeto "y" não possua nomes de colunas.
<code>precision</code>	diferença máxima aceitável entre os resíduos de duas iterações do algoritmo
<code>x, object</code>	objeto da classe <code>zigzag</code>
<code>dots</code>	parâmetros opcionais repassados para as funções de baixo nível

Detail

Esta versão do algoritmo somente aceita o caso completo. Em desenvolvimentos futuros, iremos trabalhar com outros desenhos experimentais.

Value

Um objeto da classe `zigzag` contendo os valores dos coeficientes alfa e beta além dos índices ambientais.

Examples

```
> colnames(cultivares)
> # Executando o algoritmo zigzag e salvando o resultado
em "z"
> z<-zigzag(cultivares)
> print(z)
> # NOTE que a ordem dos cultivares foi alterada!
> # Acessando as ordenadas e declives calculados
> print(z$alpha)
> print(z$beta)
```

R2***Diagnóstico do ajuste zigzag***

Description

Realiza um diagnóstico do ajuste do modelo zigzag. Veja detalhes em 3.3.3- Análise dos Resíduos.

Um valor próximo a zero indica um bom ajuste, enquanto valores maiores indicam que o algoritmo zigzag não foi capaz de extrair toda a informação relevante dos dados.

Usage

```
> R2(z)
```

Arguments

```
> z          objeto da classe "zigzag"
```

Examples

```
> z <- zigzag(cultivares)
> print(R2(z))
```

contorno

Algoritmo de contorno

Description

Realiza a análise gráfica do contorno superior, aplicado a seleção de cultivares.

Necessita de um ajuste de regressão como entrada, que pode ser gerado usando a função `\code{\link{zigzag}}`

Usage

```
> contorno(a, b = NULL, x1 = NULL, x2 = NULL, nomes =
NULL)

> method{plot}{contorno}(x, ...)

> method{print}{contorno}(x, ...)
```

Arguments

<code>a</code>	vetor com coeficientes alfa de regressão. Alternativamente, objeto da classe <code>zigzag</code> contendo os coeficientes de regressão alfa e beta e índices ambientais (ver exemplos).
<code>b</code>	vetor com coeficientes beta da regressão.
<code>x1</code>	índice ambiental mínimo.
<code>x2</code>	índice ambiental máximo.
<code>nomes</code>	vetor contendo os nomes dos cultivares.
<code>x</code>	objeto da classe <code>zigzag</code>
<code>...</code>	parâmetros opcionais repassados para as funções de baixo nível}
<code>contorno</code>	objeto da classe "contorno"

Value

Um objeto da classe "contorno"

See Also

zigzag

Examples

```

> z<-zigzag(cultivares)
> con1 <- contorno(z)
> print(con1)
> plot(con1)

# Para dados de regressão não ajustados pelo zigzag:
> conOliv<-contorno(Oliveira97$Ordenada,Oliveira97$Declive
, 0, 5000, Oliveira97$cultivar)
> print(conOliv)
> plot(conOliv)

```

compMult***Comparações Múltiplas***

Description

A função `compMult` realiza testes estatísticos de comparações múltiplas, segundo o método selecionado, para determinar quantos cultivares são significativamente dominados em cada intervalo.

Usage

```

>compMult(contorno,metodo=c("Student","Scheffe","Bonferro
ni","Tukey"), nivel = 0.05)

> Student(a, b, j, jl, mint, maxt, x, y)
> Bonferroni(a, b, j, jl, mint, maxt, x, y)
> Scheffe(a, b, j, jl, mint, maxt, x, y)
> Tukey(a, b, j, jl, mint, maxt, x, y)

```

Arguments

Contorno	objeto gerado pela função contorno
Metodo	método de comparação a ser utilizado. Pode ser escolhido entre Student (teste t), Scheffé, Bonferroni e Tukey. Veja as referências para determinar qual teste é adequado para cada caso.
Nivel	nível de significância. O valor padrão corresponde a p-value=0.05
a	vetor com coeficientes alfa da regressão.
b	vetor com coeficientes beta da regressão.
j, j1	índices do vetor a serem comparados
mint, maxt	intervalo de comparação
x	índices ambientais
y	matriz de produtividade dos cultivares

Detail

As funções `Student()`, `Bonferroni()`, `Scheffe()` e `Tukey()` são funções auxiliares utilizadas para retornar o valor p do teste associado. Normalmente, nenhuma destas deve ser usada pelo utilizador final.

A função `contorno` pode ser invocada a partir de um objeto `zigzag` ou a partir de coeficientes de regressão obtidos de outra forma. Neste último caso, não é possível realizar as comparações múltiplas, pois os dados de produção originais são necessários.

Examples

```
> z <- zigzag(cultivares)
> cont <- contorno(z)
> compMult (cont, nivel=0.05)
```

5.2. Estrutura do *package*

```

#' Algoritmo zigzag

> zigzag <- function(y, nomes=NULL, precision=1e-6) {
+   n <- dim(y)[2]
+   if(is.null(nomes)){
+     if(is.null(colnames(y)))
+       stop("Deve ser fornecido um vetor de nomes para os
cultivares")
+     nomes <- colnames(y)
+   }
+   if(length(nomes) != n) stop ("O numero de nomes deve ser
igual ao numero de colunas em y")
+   # chute inicial para x_i: rendimentos médios dos blocos
+   x <- apply(y, 1, mean)
+   sumres <- Inf
+   while(TRUE){
+     # geramos alfa e beta a partir dos mínimos quadrados
+     b <- apply(y, 2, get_b, x=x)
+     a <- c()
+     for(i in 1:length(b))
+       a <- c(a, get_a(y[,i], x=x, b=b[i]))
+     # intermediários para gerar novo x
+     newy <- y
+     for (i in 1:length(b))
+       newy[,i] <- b[i]*y[,i]
+     newx <- (apply( newy, 1, sum) - sum(a*b)) / sum(b*b)
+     # Reescalando
+     m0 = min(x); M0 = max(x);
+     m1 = min(newx); M1 = max(newx);
+     newx <- m0 + (M0-m0)/(M1-m1) * (newx - m1)

```

CAPÍTULO 5

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```
+ # Calcula a nova soma de resíduos quadrados e compara
com a anterior.
+ newsumres <- sum(apply(calc.residuals(y, newx, a, b)^2,
2, sum))
+ if (sumres - newsumres < precision) break;
+ # Se chegou aqui, retoma o loop com os novos valores de x
+ x <- newx
+ sumres <- newsumres
+ }
+ # Calcula os valores de R2 para cada cultivar:
+ resid2 <- calc.residuals(y, x, a, b)^2
+ resid2 <- apply(resid2, 2, sum)
+ f <- function(r) sum((r -mean(r))^2)
+ tot2 <- apply(y, 2, f)
+ R2 <- 1 - resid2/tot2
+ # Ordena os cultivares pela ordem de b
+ ordem <- order(-b); a <- a[ordem]; b <- b[ordem]; y <-
y[, ordem]; nomes <- nomes[ordem]
+ R2 <- R2[ordem]
+ resultado <- list(alpha=a, beta=b, x=x, y=y, nomes=nomes,
R2=R2)
+ class(resultado) <- "zigzag"
+ return(resultado)
+ }
> #' method residuals zigzag
> residuals.zigzag <- function(object, ...) {
+ calc.residuals(object$y,object$x,object$alpha,object$beta)
+ }
> # Função interna, usada em residuals como em zigzag
> calc.residuals <- function(y, x, a, b) {
+ cols = dim(y)[2]
+ x <- matrix(rep(x, cols), ncol=cols)
+ x <- sweep(x, MARGIN=2, b, '*')
```

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```

+   x <- sweep(x, MARGIN=2, a, '+')
+   y - x
+ }
> #' Order zigzag
> print.zigzag <- function(x, ...) {
+   cat("Resultado do algoritmo zigzag\n")
+   zzd <- data.frame(Cultivar = x$nomes, Ordenada = x$alpha,
Declive = x$beta, R2 = x$R2)
+   rownames(zzd) <- NULL
+   print(zzd)
+   cat("\nÍndices ambientais (x):\n")
+   print(x$x)
+   cat("\n Média dos Índices ambientais (x):\n")
+   media=mean(x$x)
+   print(media)
+   cat("\nR2 diagnóstico (deve ser próximo de zero):\n")
+   print(R2(x))
+ }
> # Funções auxiliares para cálculo dos mínimos quadrados.
Retornam alfa e beta
> get_a <- function(y, x, b) mean(y) - b*mean(x)
> get_b <- function(y, x) {
+   ybar <- mean(y);
+   xbar <- mean(x);
+   n <- sum( (x-xbar)*(y-ybar))
+   d <- sum((x - xbar)^2)

##' Diagnóstico do ajuste zigzag

> R2 <- function(z) {
+   y <- unlist(residuals(z))
+   SSzigzag <- t(y) %*% y
+   Xb <- matrix(c(rep(1, length(z$x)), z$x), ncol=2)
+   X1 <- diag(length(z$beta)) %x% Xb

```

CAPÍTULO 5

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```
+ X2 <- z$beta %x% diag(length(z$x))
+ Xbb <- cbind(X1, X2)
+ SSmodelo <- SSzigzag - t(y) %*% Xbb %*% ginv(t(Xbb) %*%
Xbb) %*% t(Xbb) %*% y
+ cat("SS zigzag:",SSzigzag," SS modelo:", SSmodelo, "\n")
+ return(as.numeric(1 - SSmodelo/SSzigzag))
+ }
```

#' Algoritmo de contorno

```
> contorno<-function(a, b = NULL, x1 = NULL, x2 = NULL, nomes
= NULL) {
+   ret.x <- c(x1,x2); ret.y <- NULL
+   # Extrai os dados de alfa, beta, x1 e x2 do objeto "a"
+   if(class(a) == "zigzag") {
+     b = a$beta; x1 = min(a$x); x2 = max(a$x); nomes=a$nomes
+     ret.x <- a$x
+     ret.y <- a$y
+     a = a$alpha
+   }
+   if(length(a)!=length(b)) #teste se os coeficientes de
regressão ajustados têm o mesmo tamanho.
+   {
+ stop("Os coeficientes de regressão "a" e "b" (alfa e beta)
devem possuir o mesmo número de observações")
+   }
+   n<-length(b) #atribui a "n" o número de declives b
+   if (! identical(order(-b), 1:n))
+ warning("Aviso: cultivares reordenados pela ordem de maior
declive")
+   nomes <- nomes[order(-b)]
+   va<-a[order(-b)]
+   vb<-b[order(-b)]
+   # Comparação de cultivares
```

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```

+ tetas <- c()
+ dominancia <- c()
+ # Determina qual cultivar tem maior valor em x2
+ u <- which((x2 * vb + va) == max(x2 * vb + va))
+ xmax <- x2
+ while(xmax>x1) #comparação entre cultivares até que o
índice máximo seja menor que o índice mínimo
+ {# Calcula as interseções entre as retas
+ intercept <- (va[u] - va) / (vb - vb[u])
+ # Remove valores inválidos (NA, infinito, superior a x2)
+ intercept [is.infinite(intercept)]<-0
+ intercept [is.na(intercept)]<-0
+ for (j in 1:n)
+   if (intercept [j]>=xmax)
+     intercept [j]<-0
+ # Nosso valor de teta é a maior interseção permitida
+ teta<-max(intercept)
+ # Salva as informações para serem mostradas
posteriormente
+ tetas <- c(tetas, teta)
+ dominancia <- c(dominancia, u)
+ xmax<-teta # o índice máximo passa a tomar o valor da
variável "teta"
+ # Escolhe o próximo intercepto
+ for(j in 1:n) {
+   if (intercept [u]< intercept [j])
+     u<-j
+   else if(intercept[u]== intercept [j] & va[u]+vb[u]*(
intercept [u])<va[j]+vb[j]*( intercept[j]))
+     u<-j
+   }
+ }
+ # Ultimo teta assume valor de x1

```

CAPÍTULO 5

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```
+ # Isso será usado no algoritmo de compMult
+ tetas[length(tetas)] <- x1
+ result <- list(a = va, b = vb, x = ret.x, y=ret.y, teta =
tetas, dominancia = dominancia, nomes=nomes)
+ class(result) <- "contorno"
+ return(result)
+ }
> # Importação de gráficos
> plot.contorno <- function(x, ...) {
+ dots <- list(...)
+ a <- x$a; b <- x$b; x1 <- min(x$x); x2 <- max(x$x);
+ teta <- x$teta; dominancia <- x$dominancia; nomes <-
x$nomes
+ if (is.factor(nomes))
+   nomes <- levels(nomes)
+ # Se não foi informada paleta, usa a paleta escolhida
como padrão
+ if (! "col" %in% names(dots))
+   col <-c("blue","red","yellow","cyan","darkred","orange",
"chartreuse2","darkgoldenrod3","burlywood","darkcyan",
"darkmagenta","darkgrey","pink","blueviolet")
+ else col <- dots$col
+ if(!"main" %in%names(dots))dots$main<-"Contorno superior"
+ if (!"xlab" %in% names(dots))dots$xlab="Índice Ambiental"
+   if (! "ylab" %in% names(dots)) dots$ylab="Produção
(kg/ha) "
+ # Encontra o valor máximo a ser usado no eixo "y":
+ maxp <- max(c(max(x1 * b + a), max(x2 * b + a)))
+ # Inicializa o gráfico
+ do.call(plot, c(list(x=0,y=0, bty="l", type='n', xlim=c(0,
x2), ylim=c(0, maxp)), dots))
+ # Cria o looping que desenha o gráfico com os diversos
cultivares devidamente legendados#
```

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```

+   for (i in 1:length(b)) {
+       abline(a[i],b[i],col=col[i]) #insere "n" retas no
referencial, de cores diferentes
+   }
+   abline(v=x1,col="darkgrey",lty=2)      #Plota uma linha
tracejada vertical no índice x1
+   text(x1,0,"Tm",cex=0.87, adj=c(0,1)) #Atribui à linha
vertical x1 o nome "índice mínimo"
+   abline(v=x2,col="darkred",lty=2)      #Plota uma linha
tracejada vertical no índice x2
+   text(x2,0,"TM",cex=0.87, adj=c(1,1)) #Atribui à linha
vertical x2 o nome "índice máximo"
+   for (i in 1:length(nomes))
+       if(i %in% dominancia)
+           nomes[i] <- paste0(nomes[i], '*')
+
legend("topleft",legend=nomes,cex=0.9,bty="n",ncol=2,fill=col
)
+   legend("bottom",paste("Contorno Superior"),cex=0.7,lwd=2)
#coloca uma legenda a indicar o Contorno Superior
+   # Traça em verde as linhas correspondentes aos valores de
teta:
+   oldt <- x2 # na primeira iteração, o valor "até onde for"
é o valor máximo
+   for (i in 1:length(teta)) {
+       if (teta[i] > x1) {
+           abline(v=teta[i],col=3,lty=2) #Plota linhas verticais
verde, nas interseções
+           text(teta[i],0,paste0("T",i),cex=0.87,      adj=c(0,1))
#atribui à linha vertical o nome "teta i"
+           # O programa irá colocar no gráfico o contorno superior
destacado
+           y<-a[dominancia[i]]+b[dominancia[i]]* c(teta[i], oldt)

```

CAPÍTULO 5

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```
+     lines(c(teta[i], oldt), y, type="l",lwd=2)
+   } else {
+   # Primeiro cultivar do contorno. Não traça linha vertical
e inicia em x1
+     y<-a[dominancia[i]]+b[dominancia[i]]* c(x1, oldt)
+     lines(c(x1, oldt), y, type="l",lwd=2)
+   }
+   oldt <- teta[i]
+ }
+ }
> #' Order contorno
> print.contorno <- function(x, ...) {
+   cat("Análise de contorno superior")
+   a <- x$a; b <- x$b; x1 <- min(x$x); x2 <- max(x$x);
+   teta <- x$teta; dominancia <- x$dominancia; nomes <-
x$nomes
+   oldt <- x2 # Na primeira iteração, o valor "até onde for"
+   é o valor máximo
+   for (i in 1:length(teta)) {
+     if (teta[i] > x1) {
+       cat('\n O cultivar', nomes[dominancia[i]], 'é dominante
entre', 'T=',teta[i], 'e', 'T=',oldt)
+     } else {
+       cat('\n O cultivar', nomes[dominancia[i]], 'é dominante
entre', 'T=',x1, 'e', 'T=',oldt)
+     }
+     oldt <- teta[i]
+   }
+   if (length(teta)<length(b)) cat('\n Os restantes cultivares
não fazem parte do contorno superior!\n')
+ }
```

#' Comparações múltiplas

```

> compMult<-function(contorno,metodo=c("Student","Scheffe",
"Bonferroni","Tukey"),nivel=0.05) {
+   if(!class(contorno)=="contorno") stop("Objeto 'contorno'
deve ser da classe contorno")
+   if(is.null(contorno$y)) stop("O objeto selecionado não
possui dados de produção")
+   metodo <- match.arg(metodo)
+   a<-contorno$a;   b<-contorno$b;   x<-contorno$x;   x1<-
min(contorno$x); x2 <- max(contorno$x)
+   n <- length(b)
+   cat("Comparação múltipla usando o método de",metodo,"ao
nível de significância de",nivel*100,"%\n")
+   metodoF <- match.fun(metodo)
+   nivelB <- nivel # Para Bonferroni, o nivel vai mudar
+   # Percorre a lista de cultivares dominantes
+   # EM REVERSO
+   for (i in length(contorno$teta):1) {
+     dominados <- 0
+     mint<-contorno$teta[i]
+     if (i == 1)
+       maxt <- x2
+     else
+       maxt<-contorno$teta[i-1]
+     cat("\nIntervalo de t=", mint, " a t=", maxt, ":\n")
+     j1 <- contorno$dominancia[i]
+     for (j in 1:n) {
+       prob <- metodoF (a, b, j, j1, mint, maxt, x,
contorno$y)
+       if (metodo == "Bonferroni") {
+         if (j < j1) nivelB <- nivel / (2*(j1-1))
+         else nivelB <- nivel / (2*(n-j1))

```

CAPÍTULO 5

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```
+     }
+     if ( probab < nivelB) {
+       print(paste0("O cultivar", contorno$nomes[j]," é
significativamente dominado por ",contorno$nomes[j1],"(p =
",round(prob,3),")"))
+       dominados <- dominados + 1
+     }
+   }
+   print(paste0("Cultivares significativamente dominados: ",
dominados))
+ }
+ }
> # FUNÇÕES AUXILIARES
> #' Order compMult
> Student <- function(a, b, j, j1, mint, maxt, x, y ) {
+   # Variáveis auxiliares
+   # Determina qual teta corresponde ao teta0
+   if (j > j1)
+     teta <- mint
+   else
+     teta = maxt
+   # Graus de liberdade
+   g <- (length(x) - 2) * length(b)
+   # S = função objetivo
+   S <- sum(apply(calc.residuals(y, x, a, b)^2, 2, sum))
+   W <- 1/(x%*%x)
+   k = W * (1+teta^2)
+   t <- ((a[j] + b[j]*teta) - (a[j1] + b[j1] * teta))/sqrt(
S/g * 2*k)
+   # Realiza um teste t uni-caudal(t é SEMPRE <0)
+   return(pt(t, g))
+ }
```

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```

> #' @rdname compMult
> Bonferroni<-Student # Usam a mesma estatística
> #' Order compMult
> Scheffe<-function(a, b, j, j1, mint, maxt, x, y) {
+   if (j > j1){
+     teta <- mint
+     gl1<-length(b)-j1
+   }
+   else {
+     teta = maxt
+     gl1 <- j1-1
+   }
+   # Graus de liberdade
+   g <- (length(x) - 2) * length(b)
+   # S = função objetivo
+   S <- sum(apply(calc.residuals(y, x, a, b)^2, 2, sum))
+   W <- 1/(x%*%x)
+   k = W * (1+teta^2)
+   f <- abs(((a[j] + b[j]*teta) - (a[j1] + b[j1] * teta))) /
sqrt( S/g * 2*k * gl1)
+   # Realiza um teste F uni-caudal
+   prob <- pf(f, gl1, g, lower.tail=FALSE)
+   if (is.nan(prob)) return (1)
+   return(prob)
+ }
> # Order compMult
> Tukey <- function(a, b, j, j1, mint, maxt, x, y) {
+   if (j > j1) {
+     teta <- mint
+     gl1 <- length(b) - j1 + 1
+   }
+   else {
+     teta = maxt

```

CAPÍTULO 5

PACKAGE EM LINGUAGEM R PARA EXECUÇÃO AUTOMÁTICA DA ACR

```
+   gl1 <- jl
+ }
+ # Graus de liberdade
+ g <- (length(x) - 2) * length(b)
+ # S = função objetivo
+ S <- sum(apply(calc.residuals(y, x, a, b)^2, 2, sum))
+ W <- 1/(x%*%x)
+ k = W * (1+teta^2)
+   t <- abs(((a[j] + b[j]*teta)-(a[j1] + b[j1] *
teta)))/sqrt(S/g * k)
+ # Realiza um teste T de Tukey
+ prob <- ptukey(t,nmeans=gl1,df=g,lower.tail=FALSE)
+ if (is.nan(prob)) return (1)
+ return(prob)
```

**CAPITULO 6 - APLICAÇÃO PRÁTICA E RESULTADOS
COMPUTACIONAIS**

6. Aplicação prática e Resultados computacionais

O objetivo deste capítulo é mostrar os resultados da ACR, incluindo o cálculo dos índices ambientais, o R^2 utilizando o algoritmo zigzag do capítulo 3 e todos os testes de comparação múltipla apresentados no capítulo 4, sejam obtidos utilizando o *software* R, conforme *package* criado e descrito no Capítulo 5.

6.1. Aplicação prática

Nesta aplicação pretendemos aplicar as técnicas apresentadas nos capítulos anteriores usando para tal dados simulados. Esses dados simulados referentes a produção (Kg/hectare) de cultivares encontram-se em suporte informático no apêndice B, e considerou-se que são referentes aos anos fictícios de 2010, 2011 e 2012, nos quais foram ensaiados 16 cultivares, contendo cada bloco 11 desses cultivares conduzidos em 6 estações experimentais com a seguinte codificação:

VS, SA, SN, AMI, GOF e VISTA

A seguir, apresentamos na Tabela 6.1 a lista de cultivares ensaiados. Na Tabela 6.2 os dados estão organizados considerando uma linha por cultivar, e uma coluna para cada ano. A presença [ausência] dum cultivar, num dado ano, é indicada respetivamente por 1[0].

Tabela 6.1: Lista de cultivares ensaiados

AI901	AI903	RCRUZ	TS905
AI902	AI904	TS904	ROSA
TS901	MVER	BRAVO	F.Frances
TS902	TS903	AI905	SLuz

Tabela 6.2: Presença e ausência de cultivares nos vários anos

	2010	2011	2012
AI901	1	1	1
AI902	1	1	1
TS901	1	1	1
TS902	1	1	1
AI903	1	0	0
AI904	1	1	1
MVER	1	1	0
TS903	1	1	0
RCRUZ	1	1	0
TS904	1	0	0
BRAVO	1	1	1
AI905	0	1	1
TS905	0	1	1
ROSA	0	0	1
F.Frances	0	0	1
SLuz	0	0	1

6.1.1. Algoritmo zigzag

Aplicando o algoritmo zigzag para cada ano e cada bloco, usando a função objetivo e análise de resíduos, obtemos os coeficientes de regressão e os índices ambientais ajustados.

Esses resultados estão na Tabela 6.3 e Tabela 6.4.

Tabela 6.3: Coeficientes de Regressão ajustados e os valores de R^2

Ano	2010			2011			2012		
Cultivar	$\tilde{\alpha}$	$\tilde{\beta}$	R^2	$\tilde{\alpha}$	$\tilde{\beta}$	R^2	$\tilde{\alpha}$	$\tilde{\beta}$	R^2
AI901	-69.54	0.919	0.582	203.10	0.763	0.363	159.29	0.890	0.836
AI902	234.18	0.682	0.733	146.77	0.857	0.510	557.58	0.763	0.778
TS901	-186.82	1.457	0.889	-401.86	1.428	0.637	49.50	0.903	0.800
TS902	220.02	1.089	0.664	-248.11	1.174	0.805	460.38	0.769	0.617
AI903	-14.14	1.103	0.865						
AI904	-111.65	1.010	0.810	-181.23	0.850	0.608	-301.78	1.236	0.886
MVER	-356.44	1.179	0.931	486.37	0.645	0.324			
TS903	374.30	0.883	0.757	391.82	0.650	0.293			
RCRUZ	-42.27	0.943	0.917	-210.23	1.108	0.669			
TS904	-32.56	0.891	0.880						
BRAVO	-40.32	0.891	0.891	265.55	1.024	0.575	-263.95	1.150	0.852
AI905				-381.10	1.132	0.620	-93.20	0.984	0.857
TS905				-197.68	1.290	0.723	125.26	0.974	0.912
ROSA							142.90	0.965	0.920
F.Frances							-372.04	1.187	0.972
SLuz							-255.43	1.090	0.897

Tabela 6.4: Índices ambientais e média anual.

Ano	2010	2011	2012
Número de blocos	12	16	16
		1613.25	3051.00
		1550.81	3718.53
	1916.97	1474.10	3796.22
	2060.07	1627.18	3883.07
	2174.81	1502.70	3522.25
	2379.17	1608.13	3275.59
	813.82	1499.63	3911.90
	995.19	1484.74	3827.17
	1081.64	1476.73	3313.33
	940.32	1619.93	2984.14
	1458.27	1586.99	2642.67
	2110.03	1058.92	3488.28
	2540.41	562.54	642.54
	2836.37	817.07	848.72
		618.99	849.72
		683.65	909.18
Média	1775.588	1299.089	2791.524

Seguidamente apresenta-se a soma dos quadrados dos resíduos obtidos pelo zigzag e pelo modelo e ainda os coeficientes R^2 conforme 3.3.3 (modelo para os desvios). Os resultados obtidos evidenciam uma validação do algoritmo ou seja conseguiu-se um bom ajustamento do modelo.

Contrariamente ao que é usual, mas de acordo com a definição proposta para R^2 , obtiveram-se valores considerados baixos, o que pressupõe um bom ajustamento do

modelo e que o algoritmo zigzag foi capaz de extrair toda a informação relevante dos dados.

Tabela 6.5: Somas de quadrados dos resíduos e coeficiente R^2

	2010	2011	2012
Soma de quadrados (zigzag)	$1.4263 \cdot 10^7$	$2.0133 \cdot 10^7$	$4.1379 \cdot 10^7$
Soma de quadrados (modelo)	$1.4263 \cdot 10^7$	$2.0133 \cdot 10^7$	$4.1379 \cdot 10^7$
R^2	$2.2204 \cdot 10^{-16}$	$2.9976 \cdot 10^{-15}$	$\cong 0$

Tabela 6.6: Médias dos índices ambientais e coeficientes R^2 por bloco

Ano, local	Índice ambiental	R^2
2010, VS	2186.34	$2.6645 \cdot 10^{-15}$
2010, SA	911.25	$2.9532 \cdot 10^{-14}$
2010, SN	2250.27	$1.0325 \cdot 10^{-14}$
2011, VS	1551.94	$1.1213 \cdot 10^{-14}$
2011, SA	1544.39	$1.3067 \cdot 10^{-13}$
2011, AMI	1318.57	$1.4543 \cdot 10^{-14}$
2011, GOF	672.75	$2.186029 \cdot 10^{-13}$
2012, VS	3591.88	$5.5511 \cdot 10^{-15}$
2012, VISTA	3601.77	$7.4384 \cdot 10^{-15}$
2012, AMI	2994.86	$1.3544 \cdot 10^{-15}$
2012, GOF	841.23	$5.3068 \cdot 10^{-14}$

6.1.2. Análise do contorno superior

A análise dos contornos superiores, obtidos através dos ajustamentos ano a ano, permite-nos avaliar como foi sendo conduzido o estudo. Observe-se que, a separação dos cultivares em dominantes e dominados, não depende de se aplicar ou não a estandardização nos índices ambientais. Os gráficos dos contornos superiores para cada ano encontram-se em 6.2. Resultados computacionais). Com base nos mesmos obtivemos a Tabela 6.7. 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 6.7: Composição dos contornos superiores ano a ano

Cultivar	2010	2011	2012
AI901	P	P	P
AI902	P	P	E
TS901	D	P	P
TS902	E	P	P
AI903	P		
AI904	P	P	D
MVER	P	E	
TS903	P	P	
RCRUZ	P	P	
TS904	P		
BRAVO	P	D	P
AI905		P	P
TS905		P	P
ROSA			P
F. Frances			P
SLuz			P

6.1.3. Comparações múltiplas ao nível de significância 5%

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 proceder a realização dos testes com índices de **Student**, **Scheffé**, **Bonferroni** e de **Tukey** ao nível de significância de 5% de acordo com o estabelecido em 4.1.2. obtiveram-se os seguintes resultados:

Tabela 6.8: Cultivares dominantes e dominados: resultados da aplicação do teste t de **Student** ao nível de significância de 5%

Cultivar dominante	Intervalo de dominância	Cultivares significativamente dominados	Número de cultivares dominados
2010			
TS901	[1107.08; 2836.36]	MVER, AI903, AI904, RCRUZ, AI901, TS904, BRAVO, AI902	8
TS902	[813.818; 1107.08]	MVER, AI903, AI904, RCRUZ, AI901, TS904, BRAVO, AI902	8
2011			
BRAVO	[582.975; 1627.182]	TS902, AI905, RCRUZ, AI902, AI904, AI901, TS903	7
MVER	[562.5455; 582.975]	TS901, TS905, TS902, AI905, RCRUZ, AI902, AI904, AI901, TS903	9
2012			
AI904	[1817.307; 3911.909]	Sluz, AI905, TS901	3
AI902	[642.5455; 1817.307]	Sluz, AI905, TS901	3

Tabela 6.9: Cultivares dominantes e dominados: resultados da aplicação do método de **Scheffé** ao nível de significância de 5%.

Cultivar dominante	Intervalo de dominância	Cultivares significativamente dominados	Número de cultivares dominados
2010			
TS901	[1107.08; 2836.36]	-	0
TS902	[813.818; 1107.08]	MVER, AI904, RCRUZ, AI901, TS904, BRAVO	6
2011			
BRAVO	[582.975; 1627.182]	AI904	1
MVER	[562.545; 582.975]	TS901, TS902, AI905, RCRUZ, AI904	5
2012			
AI904	[1817.307; 3911.909]	-	0
AI902	[642.545; 1817.307]	-	0

Tabela 6.10: Cultivares dominantes e significativamente dominados: resultados da aplicação do método de **Bonferroni** ao nível de significância de 5%

Cultivar dominante	Intervalo de dominância	Cultivares significativamente dominados	Número de cultivares dominados
2010			
TS901	[1107.08; 2836.36]	MVER, AI904 RCRUZ, AI901, TS904, BRAVO, AI902	7
TS902	[813.818; 1107.08]	MVER, AI903, AI904 RCRUZ, AI901, TS904, BRAVO, AI902	8
2011			
BRAVO	[582.975; 1627.182]	AI905, AI902, AI904, AI901	4
MVER	[562.5455; 582.975]	TS901, TS905, TS902, AI905, RCRUZ, AI902, AI904, AI901, MVER	9
2012			
AI904	[1817.307; 3911.90]	-	0
AI902	[642.545; 1817.307]	AI902	1

Tabela 6.11: Cultivares dominantes e significativamente dominados: resultados da aplicação do método de **Tukey** ao nível de significância de 5%

Cultivar dominante	Intervalo de dominância	Cultivares significativamente dominados	Número de cultivares dominados
2010			
TS901	[1107.08; 2836.36]	MVER, AI904 RCRUZ, AI901, TS904, BRAVO, AI902	7
TS902	[813.818; 1107.08]	MVER, AI903, AI904 RCRUZ, AI901, TS904, BRAVO, AI902	7
2011			
BRAVO	[582.975; 1627.182]	AI905, AI902, AI904, AI901	4
MVER	[562.5455; 582.975]	TS901, TS905, TS902, AI905, RCRUZ, AI902, AI904, AI901	8
2012			
AI904	[1817.307; 3911.909]	-	0
AI902	[642.545; 1817.307]	-	0

6.2. Resultados computacionais

De seguida apresentamos os resultados referentes aos dados simulados de cultivares ensaiados nos anos 2010 a 2012, após aplicação do algoritmo zigzag, contorno superior, e comparações múltiplas.

Os dados de entrada das tabelas do apêndice B, guardados em ficheiro no formato .txt, foram importados através da função `read.table()`.

Para os diferentes anos, tendo em vista a obtenção dos resultados em simultâneo para o algoritmo zigzag, contorno superior, e comparações múltiplas usaremos o seguinte código:

```
> cultivares <- read.table("cultivaresano.txt", head=T)
  #ano= 2010, 2011 ou 2012.

> library(MASS)
> z = zigzag(cultivares)
> z
> con <- contorno(z)
> plot (con)
> print (con)
> compMult (con, metodo = "Student")
> compMult (con, metodo = "Scheffe")
> compMult (con, metodo = "Bonferroni")
> compMult (con, metodo = "Tukey")
```

6.2.1. Ano de 2010

Resultados do algoritmo zigzag

	Cultivar	Ordenada	Declive	R2
1	TS901	-186.81774	1.4568796	0.8888991
2	MVER	-356.43790	1.1788235	0.9305170
3	AI903	-14.14225	1.1034237	0.8649619
4	TS902	220.01848	1.0893941	0.6641166
5	AI904	-111.65002	1.0099848	0.8097240
6	RCRUZ	-42.27156	0.9430801	0.9173918
7	AI901	-69.54134	0.9192963	0.5822097
8	TS904	-32.56467	0.8909620	0.8797770
9	BRAVO	-40.31815	0.8908701	0.8907413
10	TS903	374.29757	0.8839528	0.7572580
11	AI902	234.18052	0.6822545	0.7331458

Índices ambientais (x)

[1] 1916.964 2060.071 2174.806 2379.168 813.818 995.190

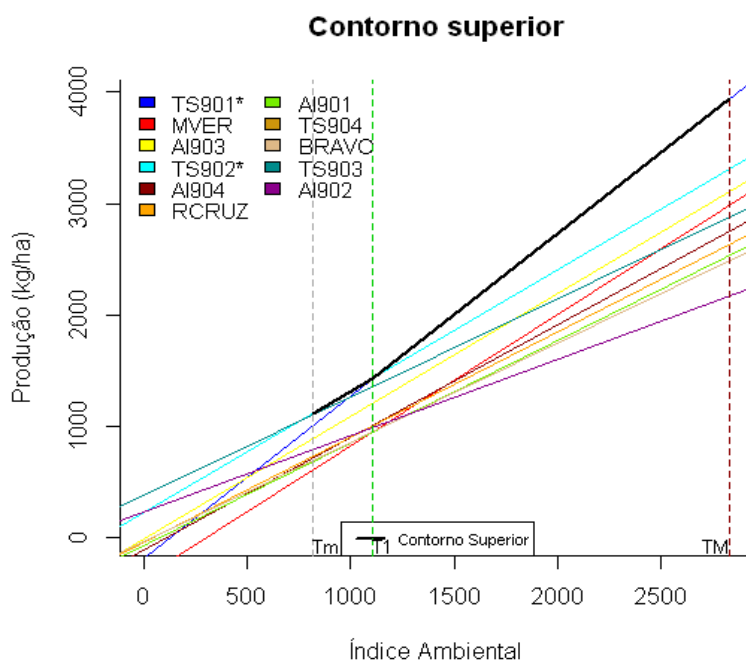
[7] 1081.641 940.320 1458.2717 2110.028 2540.408 2836.363

R2 diagnóstico (deve ser próximo de zero)

SS zigzag: 14263129 SS modelo: 14263129

[1] 2.220446e-16

Figura 6.1: Traçado do contorno superior – 2010



Análise de contorno superior

O cultivar TS901 é dominante entre $T = 1107.081$ e $T = 2836.364$

O cultivar TS902 é dominante entre $T = 813.8182$ e $T = 1107.081$

Os restantes cultivares não fazem parte do contorno superior!

Comparação múltipla usando teste t de Student ao nível de significância de 5%.

Intervalo de $t= 813.8182$ a $t= 1107.081$:

- [1] "O cultivar MVER é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar AI903 é significativamente dominado por TS902 (p = 0.006) "
- [1] "O cultivar AI904 é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar RCRUZ é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar AI901 é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar TS904 é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar BRAVO é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar AI902 é significativamente dominado por TS902 (p = 0) "
- [1] "Cultivares significativamente dominados: 8"

Intervalo de $t= 1107.081$ a $t= 2836.364$:

- [1] "O cultivar MVER é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar AI903 é significativamente dominado por TS901 (p = 0.006) "
- [1] "O cultivar AI904 é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar RCRUZ é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar AI901 é significativamente dominado por TS901 (p = 0) "

CAPÍTULO 6

APLICAÇÃO PRÁTICA E RESULTADOS COMPUTACIONAIS

[1] "O cultivar TS904 é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar BRAVO é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar AI902 é significativamente dominado por TS901
(p = 0) "

[1] "Cultivares significativamente dominados: 8"

Comparação múltipla usando o método de Scheffé ao nível de significância de 5%

Intervalo de t= 813.8182 a t= 1107.081:

[1] "O cultivar MVER é significativamente dominado por TS902
(p = 0.026) "

[1] "O cultivar AI904 é significativamente dominado por TS902
(p = 0.027) "

[1] "O cultivar RCRUZ é significativamente dominado por TS902
(p = 0.033) "

[1] "O cultivar AI901 é significativamente dominado por TS902
(p = 0.017) "

[1] "O cultivar TS904 é significativamente dominado por TS902
(p = 0.021) "

[1] "O cultivar BRAVO é significativamente dominado por TS902
(p = 0.019) "

[1] "Cultivares significativamente dominados: 6"

Intervalo de t= 1107.081 a t= 2836.364:

[1] "Cultivares significativamente dominados: 0"

Comparação múltipla usando o método de Bonferroni ao nível de significância de 5%

Intervalo de t= 813.8182 a t= 1107.081 :

[1] "O cultivar MVER é significativamente dominado por TS902
(p = 0) "

- [1] "O cultivar AI903 é significativamente dominado por TS902 (p = 0.006) "
- [1] "O cultivar AI904 é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar RCRUZ é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar AI901 é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar TS904 é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar BRAVO é significativamente dominado por TS902 (p = 0) "
- [1] "O cultivar AI902 é significativamente dominado por TS902 (p = 0) "
- [1] "Cultivares significativamente dominados: 8"

Intervalo de t= 1107.081 a t= 2836.364:

- [1] "O cultivar MVER é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar AI904 é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar RCRUZ é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar AI901 é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar TS904 é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar BRAVO é significativamente dominado por TS901 (p = 0) "
- [1] "O cultivar AI902 é significativamente dominado por TS901 (p = 0) "
- [1] "Cultivares significativamente dominados: 7"

Comparação múltipla usando o método de Tukey ao nível de significância de 5%

Intervalo de $t= 813.8182$ a $t= 1107.081$:

[1] "O cultivar MVER é significativamente dominado por TS902
(p = 0) "

[1] "O cultivar AI904 é significativamente dominado por TS902
(p = 0) "

[1] "O cultivar RCRUZ é significativamente dominado por TS902
(p = 0) "

[1] "O cultivar AI901 é significativamente dominado por TS902
(p = 0) "

[1] "O cultivar TS904 é significativamente dominado por TS902
(p = 0) "

[1] "O cultivar BRAVO é significativamente dominado por TS902
(p = 0) "

[1] "O cultivar AI902 é significativamente dominado por TS902
(p = 0) "

[1] "Cultivares significativamente dominados: 7"

Intervalo de $t= 1107.081$ a $t= 2836.364$:

[1] "O cultivar MVER é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar AI904 é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar RCRUZ é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar AI901 é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar TS904 é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar BRAVO é significativamente dominado por TS901
(p = 0) "

[1] "O cultivar AI902 é significativamente dominado por TS901 (p = 0) "

[1] "Cultivares significativamente dominados: 7"

6.2.2. Ano de 2011

Resultado do algoritmo zigzag

	Cultivar	Ordenada	Declive	R2
1	TS901	-401.8682	1.4283520	0.6371436
2	TS905	-197.6830	1.2901800	0.7233702
3	TS902	-248.1165	1.1747113	0.8052686
4	AI905	-381.1038	1.1328453	0.6201530
5	RCRUZ	-210.2360	1.1082182	0.6690115
6	BRAVO	265.5578	1.0247606	0.5754422
7	AI902	146.7752	0.8576011	0.5106673
8	AI904	-181.2399	0.8501556	0.6087978
9	AI901	203.1038	0.7631475	0.3631826
10	TS903	391.8290	0.6507320	0.2933787
11	MVER	486.3775	0.6459799	0.3248848

Índices ambientais (x)

[1] 1613.254 1550.815 1474.101 1627.181 1502.701 1608.134

[7] 1499.632 1484.748 1476.731 1619.930 1586.995 1058.920

[13] 562.545 817.0760 618.9938 683.6526

Média dos Índices ambientais (x)

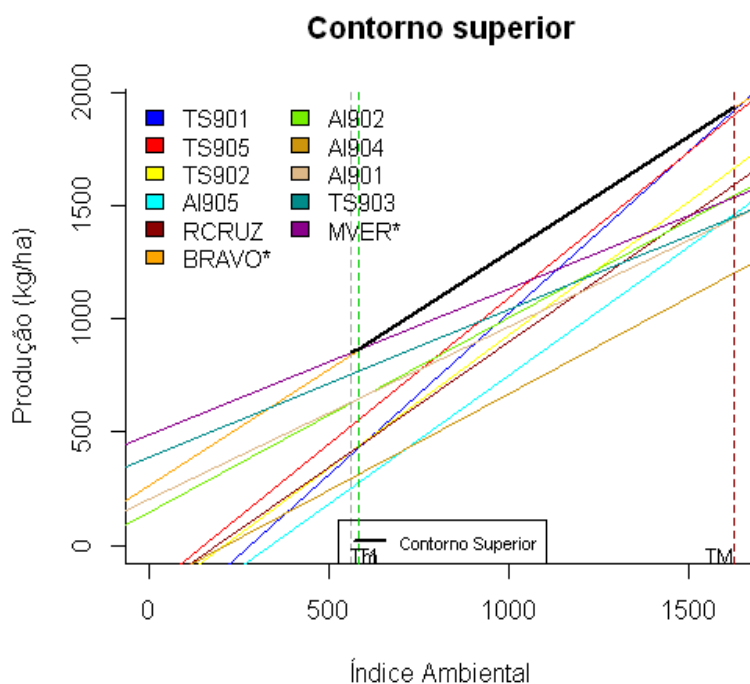
[1] 1299.089

R2 diagnóstico (deve ser próximo de zero):

SS zigzag: 20133981 SS modelo: 20133981

[1] 2.997602e-15

Figura 6.2: Traçado do contorno superior – 2011



Análise de contorno superior

O cultivar BRAVO é dominante entre $T = 582.975$ e $T = 1627.182$

O cultivar MVER é dominante entre $T = 562.5455$ e $T = 582.975$

Os restantes cultivares não fazem parte do contorno superior!

```
> compMult(con,metodo="Student")
```

Comparação múltipla usando teste t de Student ao nível de significância de 5%

Intervalo de $t = 562.5455$ a $t = 582.975$:

```
[1] "O cultivar TS901 é significativamente dominado por MVER (p = 0) "
```

```
[1] "O cultivar TS905 é significativamente dominado por MVER (p = 0) "
```

```
[1] "O cultivar TS902 é significativamente dominado por MVER (p = 0) "
```

```
[1] "O cultivar AI905 é significativamente dominado por MVER (p = 0) "
```

[1] "O cultivar RCRUZ é significativamente dominado por MVER (p = 0) "

[1] "O cultivar AI902 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar AI904 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar AI901 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar TS903 é significativamente dominado por MVER (p = 0.049) "

[1] "Cultivares significativamente dominados: 9"

Intervalo de t= 582.975 a t= 1627.182:

[1] "O cultivar TS902 é significativamente dominado por BRAVO (p = 0.04) "

[1] "O cultivar AI905 é significativamente dominado por BRAVO (p = 0.001) "

[1] "O cultivar RCRUZ é significativamente dominado por BRAVO (p = 0.014) "

[1] "O cultivar AI902 é significativamente dominado por BRAVO (p = 0) "

[1] "O cultivar AI904 é significativamente dominado por BRAVO (p = 0) "

[1] "O cultivar AI901 é significativamente dominado por BRAVO (p = 0) "

[1] "O cultivar TS903 é significativamente dominado por BRAVO (p = 0.049) "

[1] "Cultivares significativamente dominados: 7"

Comparação múltipla usando o método de Scheffé ao nível de significância de 5%

Intervalo de $t= 562.5455$ a $t= 582.975$:

[1] "O cultivar TS901 é significativamente dominado por MVER (p = 0.009) "

[1] "O cultivar TS902 é significativamente dominado por MVER (p = 0.01) "

[1] "O cultivar AI905 é significativamente dominado por MVER (p = 0.001) "

[1] "O cultivar RCRUZ é significativamente dominado por MVER (p = 0.009) "

[1] "O cultivar AI904 é significativamente dominado por MVER (p = 0.001) "

[1] "Cultivares significativamente dominados: 5"

Intervalo de $t= 582.975$ a $t= 1627.182$:

[1] "O cultivar AI904 é significativamente dominado por BRAVO (p = 0.001) "

[1] "Cultivares significativamente dominados: 1"

Comparação múltipla usando o método de Bonferroni ao nível de significância de 5%

Intervalo de $t= 562.5455$ a $t= 582.975$:

[1] "O cultivar TS901 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar TS905 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar TS902 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar AI905 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar RCRUZ é significativamente dominado por MVER (p = 0) "

[1] "O cultivar AI902 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar AI904 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar AI901 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar MVER é significativamente dominado por MVER (p = 0.5) "

[1] "Cultivares significativamente dominados: 9"

Intervalo de t= 582.975 a t= 1627.182:

[1] "O cultivar AI905 é significativamente dominado por BRAVO (p = 0.001) "

[1] "O cultivar AI902 é significativamente dominado por BRAVO (p = 0) "

[1] "O cultivar AI904 é significativamente dominado por BRAVO (p = 0) "

[1] "O cultivar AI901 é significativamente dominado por BRAVO (p = 0) "

[1] "Cultivares significativamente dominados: 4"

Comparação múltipla usando o método de Tukey ao nível de significância de 5%

Intervalo de t= 562.5455 a t= 582.975:

[1] "O cultivar TS901 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar TS905 é significativamente dominado por MVER (p = 0) "

[1] "O cultivar TS902 é significativamente dominado por MVER (p = 0) "

CAPÍTULO 6

APLICAÇÃO PRÁTICA E RESULTADOS COMPUTACIONAIS

[1] "O cultivar AI905 é significativamente dominado por MVER
(p = 0) "

[1] "O cultivar RCRUZ é significativamente dominado por MVER
(p = 0) "

[1] "O cultivar AI902 é significativamente dominado por MVER
(p = 0.006) "

[1] "O cultivar AI904 é significativamente dominado por MVER
(p = 0) "

[1] "O cultivar AI901 é significativamente dominado por MVER
(p = 0.006) "

[1] "Cultivares significativamente dominados: 8"

Intervalo de t= 582.975 a t= 1627.182:

[1] "O cultivar AI905 é significativamente dominado por BRAVO
(p = 0.03) "

[1] "O cultivar AI902 é significativamente dominado por BRAVO
(p = 0.002) "

[1] "O cultivar AI904 é significativamente dominado por BRAVO
(p = 0) "

[1] "O cultivar AI901 é significativamente dominado por BRAVO
(p = 0.002) "

[1] "Cultivares significativamente dominados: 4"

6.2.3. Ano de 2012**Resultado do algoritmo zigzag**

	Cultivar	Ordenada	Declive	R2
1	AI904	-301.78070	1.2366338	0.8863928
2	F.Frances	-372.04886	1.1873169	0.9725731
3	BRAVO	-263.95870	1.1504991	0.8523735
4	Sluz	-255.43730	1.0902421	0.8972236
5	AI905	-93.20108	0.9846596	0.8576122
6	TS905	125.26049	0.9745310	0.9129952
7	ROSA	142.90352	0.9659943	0.9206529
8	TS901	49.50888	0.9033563	0.8042343
9	AI901	159.29307	0.8900897	0.8367774
10	TS902	460.38636	0.7694904	0.6175890
11	AI902	557.58698	0.7637541	0.7782031

Índices ambientais (x):

[1] 3051.003 3718.531 3796.221 3883.078 3522.252 3275.592
 [7] 3911.909 3827.171 3313.330 2984.141 2642.676 3488.288
 [13] 642.545 848.727 849.728 909.1884

Média dos Índices ambientais (x):

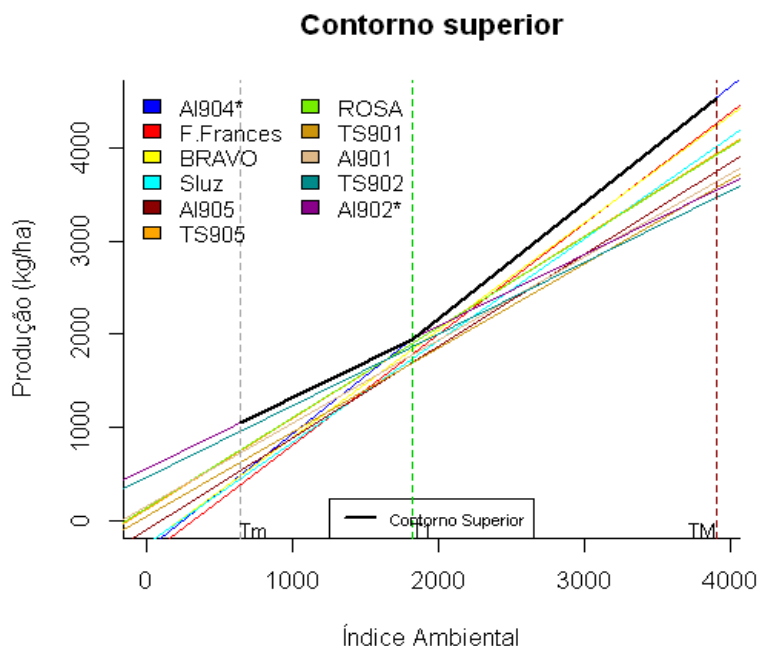
[1] 2791.524

R2 diagnóstico (deve ser próximo de zero):

SS zigzag: 41379948 SS modelo: 41379948

[1] 0

Figura 6.3: Traçado do contorno superior – 2012



Análise de contorno superior

O cultivar AI904 é dominante entre $T = 1817.307$ e $T = 3911.909$

O cultivar AI902 é dominante entre $T = 642.5455$ e $T = 1817.307$

Os restantes cultivares não fazem parte do contorno superior!

Comparação múltipla usando teste t de Student ao nível de significância de 5%

Intervalo de $t = 642.5455$ a $t = 1817.307$:

[1] "O cultivar Sluz é significativamente dominado por AI902 (p = 0.024) "

[1] "O cultivar AI905 é significativamente dominado por AI902 (p = 0.012) "

[1] "O cultivar TS901 é significativamente dominado por AI902 (p = 0.011) "

[1] "Cultivares significativamente dominados: 3"

Intervalo de $t = 1817.307$ a $t = 3911.909$:

[1] "O cultivar Sluz é significativamente dominado por AI904 (p = 0.024) "

[1] "O cultivar AI905 é significativamente dominado por AI904
(p = 0.012) "

[1] "O cultivar TS901 é significativamente dominado por AI904
(p = 0.011) "

[1] "Cultivares significativamente dominados: 3"

Comparação múltipla usando o método de Scheffé ao nível de significância de 5%

Intervalo de t= 642.5455 a t= 1817.307:

[1] "Cultivares significativamente dominados: 0"

Intervalo de t= 1817.307 a t= 3911.909:

[1] "Cultivares significativamente dominados: 0"

Comparação múltipla usando o método de Bonferroni ao nível de significância de 5%

Intervalo de t= 642.5455 a t= 1817.307:

[1] "O cultivar AI902 é significativamente dominado por AI902
(p = 0.5) "

[1] "Cultivares significativamente dominados: 1"

Intervalo de t= 1817.307 a t= 3911.909 :

[1] "Cultivares significativamente dominados: 0"

Comparação múltipla usando o método de Tukey ao nível de significância de 5%

Intervalo de t= 642.5455 a t= 1817.307:

[1] "Cultivares significativamente dominados: 0"

Intervalo de t= 1817.307 a t= 3911.909:

[1] "Cultivares significativamente dominados: 0"

CAPITULO 7- CONCLUSÕES E TRABALHOS FUTUROS

7. Conclusões e trabalhos futuros

Neste trabalho foram apresentados os principais resultados da técnica ACR para comparação e seleção de cultivares. Tendo em vista as aplicações, foram desenvolvidos resultados computacionais para criação do *package* “**jra**” em linguagem de programação R.

O *package* foi criado tendo em vista a utilização futura de profissionais e investigadores que poderão assim usufruir de uma ferramenta mais simples para tratar dados que tenham interesse em analisar. Neste caso conseguindo-se resultados mais robustos e gráficos de melhor qualidade.

No trabalho desenvolvido, abordaram-se as seguintes temáticas integrantes da técnica ACR: algoritmo zigzag, contorno superior e métodos de comparação múltipla.

Na aplicação apresentada, foi possível, através do ajustamento de regressões obter os coeficientes de regressão $(\tilde{\alpha}, \tilde{\beta})$, os índices ambientais x_i e o R^2 ajustado ano / cultivar e ano/ano nos respetivos locais de experimentação. Conseguiu-se ainda um bom ajustamento do modelo, pois, para todos os casos obteve-se os valores baixos para o R^2 contrariando o que é usual (melhor ajustamento corresponde a valores altos de R^2)

Através do método do contorno superior analisou-se os gráficos e verificou-se que o contorno superior é uma poligonal convexa constituída por segmentos pertencentes às regressões dominantes e que quando duas retas se intersectam, à direita do ponto de interseção $\theta_{j,j'}$, fica por cima a reta que tiver maior declive, e a à esquerda a que tem menor declive. Nesta análise, foi possível identificar de forma automática quais os cultivares que pertencem a esse contorno, indicando o intervalo de dominância e escrevendo que os restantes não fazem parte do contorno superior.

Através do teste t e dos métodos de comparação múltipla verificou-se quais os cultivares são significativamente dominantes ou dominados num determinado intervalo conforme Scheffé, Bonferroni e Tukey a nível de significância de 5% afim de saber quais os cultivares deverão permanecer ou serem retirados da experimentação.

Pretende-se com este trabalho ter uma base de dados capaz de analisar toda a técnica da ACR sem ter que dar trabalho extra aos investigadores. Deste modo, em trabalhos futuros

pode desenvolver-se e introduzir outras funções em R, nomeadamente para implementar a estatística F e outros procedimentos que possam contribuir para a melhoria do *package* agora criado.

No presente trabalho não se implementou no *package* a estatística F para o estudo da estabilidade dos cultivares que é traduzida pela ausência de diferenças significativas nos coeficientes angulares das regressões ajustadas, mas pode-se implementá-la com a abordagem proposta por Pinto (2006) e Oliveira (2007) uma vez que do ponto de vista da condução de planos de melhoramento é valorizada a obtenção de cultivares que tenham boa estabilidade de resposta as variações dos índices ambientais.

Embora o *package* “**jra**” funcione conforme os resultados computacionais obtidos da seção 6.2 será necessário que seja incorporado online no CRAN do R para a utilização automática dos utilizadores, pois um *package* R nada mais é do que um diretório compactado com extensão “.tar.gz”.

Para trabalho futuro e tendo em vista a transformação do nosso diretório num *package* para ser submetido ao CRAN do R deve-se:

- Estruturar o diretório conforme as normas internacionais (o nosso ainda não está totalmente conforme).
- Executar R CMD build num diretório contendo o *package*. Este comando compacta o diretório em um arquivo “tar.gz”.
- Executar R CMD check no arquivo compactado. Este comando verifica possíveis erros, inconsistências, funções sem documentação, chamadas a *packages* que não foram explicitamente carregados, etc. Executando R CMD check no nosso *package*, ele avisa que existem “caracteres não-ASCII” no código - ou seja, acentos e cedilhas, que devem ser removidos para enviar o *package* para o CRAN.
- Executar a instalação do *package* com “install.packages (“jra_0.0.1.tar.gz”, repôs = NULL)” e verificar que todas as funções funcionam como desejado.

Referências Bibliográficas

- [1] **Bazarra, M.S., Sherali, H. D., Shetty, C.M. (2006)** – *Nonlinear programming, Theory and Algorithms*. 3rd ed., Jonh Wiley & Sons - New York.
- [2] **Becker, H. C. and J. Leon (1988)** – *Stability analysis in plant breeding*. Plant Breeding. 101 : 1 - 23.
- [3] **Daniel, B. Wright and Kamala, London (2009)** – *Modern Regression Techniques Using R: A Practical Guide*. SAGE, London, UK, 2009. Disponível em: <http://www.uk.sagepub.com/booksProdSampleMaterials.nav?prodId=Book233198> [24 de dezembro de 2014]
- [4] **Dias, C. P . (2000)** – *Análise Conjunta Pesada de Regressões*. Tese de Mestrado. Universidade de Évora.
- [5] **Digby, P.G.N. (1979)** – *Modified joint regression-analysis for incomplete variety × environment data*. Journal of Agricultural Science 93: 81–86.
- [6] **Everitt, Brian. S., Hothorn, T. (2006)** – *A Handbook of Statistical Analysis using R*. Chapman & Hall/CRC.
- [7] **Faraway, J. J. (2002)** – *Practical Regression and Anova using R*, 213p. Disponível em: <https://cran.r-project.org/other-docs.html> [12 de janeiro de 2015]
- [8] **Faraway, J. J. (2004)** – *Linear Models with R*, 240 p., Chapman & Hall/CRC, EUA.
- [9] **Faraway, J. J. (2006)** – *Extending the Linear Model with R, Generalized Linear, Mixed Effets and Nonparametric Regression Models*. Chapman & Hall/CRC.
- [10] **Farnsworth, Grant V. (2008)** – *Econometrics in R*. Disponível em: <http://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf>. [14 de novembro de 2014]
- [11] **Ferreira, T., Mexia, J. T., Baeta, J., Maças, B., Gusmão, L., Bagulho, F. e Coutinho, J. (1990)** – *Comportamento relativo das cultivares de triticales produzidas na E.N.M.P., no período 1987/88 e 1988/89: II. Aferido pela análise de regressão conjunta*. Em "V Reunião Portuguesa sobre Triticale - Elvas, Maio 1990 – Melhoramento, vol. 32.
- [12] **Fisher, R.A.; Mackenzie, W.A. (1923)** – *Studies in crop variation. II. The manorial response of different potato varieties*. The Journal of Agricultural Science 13: 311-320.

- [13] **Figueiredo, Fernando., Figueiredo, Adelaide., Ramos, Alexandra., Teles, Paulo (2009)** – *Estatística Descritiva e Probabilidades: Problemas resolvidos e propostos com aplicações em R*. 2ª Edição. Escolar Editora. Lisboa.
- [14] **Finlay, K. W., Wilkinson, G. N. (1963)** – *The analysis of adaptation in a plant breeding programme*. - Aust. J. Aric. Res., 14 : 742 - 754.
- [15] **Fox, John. (2009)** – *Aspects of the Social Organization and Trajectory of the R Project*. R Journal. Disponível em: http://journal.r-project.org/archive/2009-2/RJournal_2009-2_Fox.pdf [9 de fevereiro de 2015]
- [16] **Girke, Thomas [s.d]** – *Programming in R*. Disponível em: <http://manuals.bioinformatics.ucr.edu/home/programming-in-r> [9 de outubro de 2014]
- [17] **Gusmão, L. (1985)** – *An adequate design for regression analysis of yield trials*. Theoretical and Applied Genetics 71: 314-319.
- [18] **Gusmão, L. (1986a)** – *Inadequacy of blocking in cultivar yield trials*. Theor. Appl. Genet., 72: 98 - 104.
- [19] **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.
- [20] **Gusmão, L. (1988)** – *Assessing small grain cultivars for yield in variable environments through Joint Regression Analysis*. Rachis, 7: 22 - 25.
- [21] **Gusmão, L. (1990)** – *Avaliação de cultivares (uma perspectiva experimental diferente)*. Revista de Ciências Agrárias – nº2, vol. 13: 11 - 18.
- [22] **Gusmão, L., Baeta, J. and Mexia, J. T. (1991a)** – *A avaliação da aptidão agronómica relativa em cultivares de cereais*. In “1ª. Jornadas de Estatística e Aplicações”. Universidade de Minho – Braga, 22 a 24 de Abril.
- [23] **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) Advanes in Lupin Resear h Rd. ISA, Lisboa, Pág. 409 - 412.
- [24] **Gusmão, L., Mexia, J.T., Baeta, J. (1992a)** – *Trimmed Joint Regression: a new approach to the Joint Regressional Analysis for cultivar relative performance evaluation*. Theoretical and Applied Genetics, 84 : 735 - 738.

- [25] **Gusmão, L.; Baeta, J.; Mexia J. T.; Bagulho, F.; Maças, B. e Coutinho, J.(1992b)** – *Statistical analysis of randomised complete block designs for Barley cultivar yield trials*. *Rachis*, 11 : 14 – 18.
- [26] **Gusmão, L.; Rebelo, D. C. ; Miranda, M. J.; Baeta, J. and Vaz, F. (1993a)** – *Assessment of specific genotype x environment interations in Lupinus luteusL*. *Broteria Genética*, Lisboa, 14 : 141 - 144.
- [27] **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, nº 32.
- [28] **Gusmão, L.; Fernandes, D.; Maças, B.; Mexia J. T.; and 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.
- [29] **Ihaka, Ross e Gentleman, Robert. (1996)** – *R: A language for data analysis and graphics*. *Journal of Computational and Graphical Statistics*, 5 : 299–314.
- [30] **Júnior, Paulo J. R.(2005)**–*Introdução ao Ambiente Estatístico R*. Disponível em: <http://www.leg.ufpr.br/~paulojus/embrapa/Rembrapa/> [8 de junho de 2015]
- [31] **Knell, Robert J. (2013)** – *Introductory R : A Beginner's Guide to Data Visualisation and Analysis using R*. ISBN 978-0-9575971-0-5.
Disponível em: <http://www.introductoryr.co.uk> [4 de novembro de 2014]
- [32] **Leisch, F. (2009)** – *Creating R Packages: A Tutorial*. Paula Brito, editor, *Compstat 2008 - Proceedings in Computational Statistics*. Physica Verlag, Heidelberg, Germany.
- [33] **Lin, C. S., Binns, M. R. and Lefkovitch, L. P. (1986)** – *Stability analysis: Where Do We Stand?* *Crop. Sci.*, 26: 894 - 900.
- [34] **Maças, B.; Coutinho, J. ; Bagulho, F.; Mexia, J. T.; Baeta, J.; Gusmão, L. And Ferreira, T. (1990)** – *Comportamento relativo das cultivares de triticales produzidas na E.N.M.P., no período 1987/88 e 1988/89: I. Aferido pela análise convencional*. In “V Reunião Portuguesa sobre Triticale” – Elvas, Maio 1990 – *Melhoramento*, vol. 32.
- [35] **Mexia, T., Gusmão, L., Baeta, J. (1991)** – *Analysis of cultivar yield trials designed in Randomize Complete Blocks* - *Revista de Ciências Agrárias*, 14: 59 - 63.

- [36] **Mexia, T. (1995)** – *Introdução à Inferência Estatística Linear*. Edições Lusófonas, Lisboa.
- [37] **Mexia, J. T.; Dias, G.C. e Neves, M. (1996)** – *Roots of linear regressions, extensions and applications to joint regression analysis*. Trabalhos de investigação nº1. FCT/UNL.
- [38] **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.
- [39] **Mexia, J.T., Pereira, D.G.; Baeta, J. (1999)**–*L₂ environmental indexes*. Biometrical Letters 36: 137–143.
- [40] **Mexia, J.T., Pereira, D.G., Baeta, J. (2001)** – *Joint Regressions Analysis for winter rye cultivars using L₂ index* . Colloquium Biometryczne, 31: 207-2012.
- [41] **Mexia, J.T., Pereira, D.G., Baeta, J. (2001)** – *Weighted linear joint regression analysis*. Biometrical Letters, 38: 33 - 40.
- [42] **Mexia, J.T., Pinto, I. (2003)** – *Joint regression analysis and connectedness - Application to grapevines*. Colloquium Biometryczne, 33, 235–248.
- [43] **Ng, M.P. and Grunwald, G. K. (1997)** – *Nonlinear regression analysis of the joint-regression model*. - Biometrics, 53 : 1366 - 1372.
- [44] **Ng, M.P., Williams, E.R. (2001)** – *Joint regression analysis for incomplete two-way tables*. Australian & New Zealand Journal of Statistics 43: 201–206.
- [45] **Oliveira, A. (2007)** – *Estabilidade em análise conjunta de regressões e condição dinâmica de Planos de melhoramentos*. Tese de Doutorado Universidade Aberta.
- [46] **Oliveira, A., Oliveira, T., Mexia, J.T. (2008)** – *Analysis of residuals and adjustment in JRA*. Biometrical Letters, 45(1), 45-54.
- [47] **Oliveira, A., Oliveira, T., Mejza, S., Mexia, J.T. (2008)** – *Joint regression analysis applied to genotype stability evaluation over years*. Biuletyn Instytutu Hodowlii klimatyzacji Roslin, 250, 225-235.
- [48] **Oliveira, A., Oliveira, T.A., and Mejza, S. (2012)** – *Joint Regression Analysis and AMMI model applied to oat improvement, in Proceedings of ICNAAM 2012, Kos Island - Greece, September 19-25*. AIP Conf. Proc. 1479: 1720-1723. American Institute of Physics, ISBN 978-0-7354-1089-3.
- [49] **Pereira, D.G.; Mexia, J.T. (2002)**– *Multiple comparison in joint regression analysis with special reference for variety selection*. Agriculture, 3, 67–74.

- [50] **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.
- [51] **Pereira, D.G., Mexia, J.T. (2003b)** – *Reproducibility of joint regression analysis*. Colloquium Biometryczne, 33, 279–291.
- [52] **Pereira, D.G. (2004)** – *Análise Conjunta Pesada de Regressões em Redes de Ensaio*. Tese de Doutoramento Universidade de Évora.
- [53] **Pereira, D.G.; Mexia, J.T.; Rodrigues, P.C. (2007)** – *Robustness of joint regression analysis*. Biometrical Letters 44: 105-128.
- [54] **Pereira, D., Mexia, J.T. (2008)** – *Selection proposal of cultivars of spring barley in the years from 2001 to 2004, using joint regression analysis*. Plant Breeding, 127, 452–458.
- [55] **Pereira, D.G., Mexia, J.T. (2010)** – *Comparing double minimization and zig-zag algorithms in joint regression analysis: the complete case*. Journal of Statistical Computation and Simulation 80: 133–141.
- [56] **Perkins, J.M., Jinks, J.L. (1968)** – *Environmental and genotype environmental components of variability. III Multiple lines and crosses*. Heredity, 23: 339 – 356.
- [57] **Pinto, I.M.S. (2006)** – *Análise Conjunta de Regressões e Planos de Melhoramento*. - Tese de Doutoramento Universidade Nova de Lisboa.
- [58] **R Core Team (2014)** – *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em : <http://www.R-project.org/> [15 de outubro de 2014]
- [59] **Rodrigues, P.C., Pereira, D., Mexia, J.T. (2011)** – *A comparison between JRA and AMMI: the robustness with increasing amounts of missing data*. Scientia Agricola 68: 679–686.
- [60] **Romagosa, I., Fox, P.N. (1993)** – *Genotype x environment interaction and adaptation*. In M.D. Hayward, N.O. Bosermark & I. Romagosa, eds. *Plant breeding: principles and prospects*, p. 373-390ts. Chapman and Hall. London.
- [61] **Seber, G.A.F.; Lee, A.J. (2003)** – *Linear Regression Analysis*. Wiley-Interscience, Hoboken, NJ, USA.
- [62] **Sheffé, H. (1959)** – *The Analysis of Variance*. John Willey and Sons – New York. Review disponível em:

<http://books.google.cv/books?id=h7NuoPIXh9UC&lr=&sitesec=reviews> [6 de dezembro de 2014]

- [63] **Tavares, Victor (2013)** – *Análise Conjunta de Regressões – Ajustamento e zonagem agrícola*. Tese de Doutoramento Universidade de Évora.
- [64] **Torgo, Luís (2006)** – *Introdução à Programação em R*, 99p., Faculdade de Economia, Universidade do Porto, Porto.
- [65] **Torgo, Luís (2009)** – *A linguagem R, Programação para a análise de dados*, Escolar Editora, Lisboa.
- [66] **Vance, A. (2009)** – *Data analysts captivated by R's power*. The New York Times, 158(54548). Disponível em: http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=2&pagewanted=1 [28 de fevereiro de 2015]
- [67] **Venables, W. N., Smith, D. M. and the R Core Team (2015)** – *An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics Version 3.2.1*. Disponível em: <http://cran.r-project.org/doc/manuals/R-intro.pdf> [28 de julho de 2015]
- [68] **Verzani, J. (2005)** – *Using R for Introductory Statistics*, Chapman & Hall/CRC. Disponível em: <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf> [25 de novembro de 2014]
- [69] **Westcott, B. (1986)** – *Some methods of analyzing genotype environment interaction* Heredity, 56: 243 - 253.
- [70] **Yates, F. and Cochran, W. G. (1938)** – *The analysis of groups experiments*. J. Agric. Sci., Cambridge, 28 : 556-580.

APÊNDICES

APÊNDICE A: PROGRAMA R PARA AS FIGURAS 4.1 A 4.6

Apêndice A: Programa que permite aplicar exemplos do Método do contorno superior para os gráficos 4.1 a 4.6

```

> contorno<-function(a,b,x1,x2){
+if(length(a)!=length(b)) #condicionante realiza um teste
lógico se os coeficientes de regressão ajustados têm o mesmo
tamanho.
+ {
+ stop("Os coeficientes de regressão "a" e "b" (alfa e
beta)devem possuir o mesmo número de observações") #Retorna a
mensagem de erro caso os coeficientes tenham tamanhos
diferentes.
+ }
+matriz=data.frame(a,b) # "a" e "b" vetores que representam
ordenada e declive.
+          matriz_dec_b<-(matriz[order(matriz$b,matriz$a,
decreasing=T),])
+ va<-matriz_dec_b$a #cria um vetor com as ordenadas por
ordem do declive
+ vb<-matriz_dec_b$b #cria um vetor com os declives por ordem
decrecente
+ n<-length(vb)#atribui a "n" o número de declives b
+ x<-c(x1,x2) #atribui a "x" os índices ambientais mínimo e
máximo
+          z<-c("blue","red","yellow","cyan","darkred","orange",
"chartreuse2","darkgoldenrod3","burlywood"+          ,"darkcyan",
"darkmagenta","darkgrey","pink", +"blueviolet")
+          produção<-c(-2,x[2]*vb[1]+va[1]+2)#eixo do "y" do
referencial
+indice_Amb<-c(-2,x[2]+2)#eixo do "x" do referencial
+ plot(indice_Amb, produção, bty="l",xlab="Índice
Ambiental",ylab="Produção(kg/ha)")#Plota o gráfico cartesiano

```

APÊNDICE A

```
+for (i in 1:n)#Cria o looping que desenha o gráfico com os
diversos cultivares devidamente legendados#
+ {
+   abline(va[i],vb[i],col=z[i]) #insere "n" retas no
referencial, de cores diferente
+   abline(v=x[1],col=" darkgrey ",lty=2) #Plota uma linha
tracejada vertical no índice x[1] com uma determinada cor.
+   text(x[1],0,"Tm",cex=0.87) #atribui a linha vertical x[1] o
nome "índice mínimo"
+   abline(v=x[2],col=" dark red ",lty=2) #Plota uma linha
tracejada vertical no índice x[2] com uma determinada cor.
+   text(x[2],0,"TM",cex=0.87) #atribui à linha vertical
x[2] o nome "índice máximo"
+legend("topleft",paste("cultivar",1:n),cex=0.9,bty=
+"n",ncol=2,fill=z)
+   }#com cores diferentes é inserido legenda no canto
superior esquerdo
+ #Comparação de cultivares
+ intercept<-0
+ for (i in 1:n) #Cria um loop que compara o cultivar 1 com
os restantes e será rodado um número de vezes igual a "n".
+ {
+   intercept[i]<-c((va[1]-va[i])/(vb[i]-vb[1]))
+ }#intercept[i] corresponde aos valores de interseções.
Neste caso, é calculada todas as interseções entre o cultivar
1 e os restantes cultivares.
+ intercept[is.infinite(intercept)]<-0 #todas interseções
infinitas tomam o valor zero.
+   intercept[is.na(intercept)]<-0 #todas interseções
indeterminadas tomam o valor zero.
+   teta<-max(na.exclude(intercept))#a variável "teta"
assume o máximo de todas as interseções.
```

```

+ abline(v=teta,col=3,lty=2) #Plota linhas verticais verde,
nas interseções
+ text(teta,0,paste("T",1),cex=0.87) #atribui à linha
vertical "teta" o nome "T1" (teta1)
+ #comparação da variável "teta" com os índices ambientais
mínimo e máximo e toma a decisão de acordo com as indicações
do ciclo "if"
+ if(x[1] > teta)
+ {print('O cultivar 1 é dominante')}
+ y <-va[1]+vb[1]*x
+ lines(x,y,type="l",lwd=2) #entre o índice ambiental mínimo
e máximo é colocado no gráfico o contorno superior.
+ x[2]=x[1]-1
+ } else
+ if(x[2]<=teta)
+ {print('O cultivar 1 não faz parte do contorno superior')}
+ } else
+ {cat ('\n O cultivar 1 é dominante entre', 'T1=',teta,'e',
'TM=',x[2] )
+ h<-c(teta,x[2])
+ y<-va[1]+vb[1]*h
+ lines(h,y,type="l",lwd=2) # entre o valor da variável
"teta" e o índice ambiental máximo é colocado no gráfico o
contorno superior.
+ x[2]<-teta #o índice máximo passa a tomar o valor da
variável "teta"
+ }
+ # O ciclo for abaixo, cria um loop que compara os valores
de teta no vetor "interseção", caso dois ou mais valores
forem iguais o próximo cultivar a ser selecionado é o que
tiver maior produção antes da interseção
+ t<-1
+ u<-1

```

APÊNDICE A

```
+ for(i in 1:n)
+   {if (intercept[u]<intercept[i])
+     u<-i
+   else
+ if(intercept[u]==intercept[i] & va[u]+vb[u]*(intercept[u])
<va[i]+vb[i]*(intercept[i]))
+ u<-i else u<-u
+ }
+ while(x[2]>x[1]) #comparação entre cultivares até que o
índice máximo seja menor que o índice mínimo
+ {for (j in 1:n)
+ intercept[j]<-c((va[u]-va[j])/(vb[j]-vb[u]))
+ intercept [is.infinite(intercept)]<-0
+ intercept [is.na(intercept)]<-0
+ for (j in 1:n)
+ if (intercept [j]>=x[2])
+ intercept [j]<-0
+ teta<-max(na.exclude(intercept))
+ abline(v=teta,col=3, lty=2) #Plota linhas verticais verde,
nas interseção
+ if (x[1]>teta)
+ {cat('\n O cultivar', u, 'é dominante entre', 'Tm=',x[1],
'e', 'TM=',x[2])
+ y<-va[u]+vb[u]*x
+ lines(x,y,type="l",lwd=2) #o programa irá escrever colocar
no gráfico o contorno superior destacado entre o índice
ambiental mínimo e máximo
+ t<-t+1
+   } else
+ if (x[2]<teta)
+ {cat('\n O cultivar ',u,' não faz parte de contorno
superior')
+ t<-t
```

```

+   } else
+ {cat  ('\n  O cultivar',u,'é dominante entre',  "T",
t+1,'=',teta, 'e', "T",t,'=',x[2])
+   m<-c(teta,x[2])
+   y<-va[u]+vb[u]*m
+   lines(m,y,type="l",lwd=2) #o programa irá escrever colocar
no gráfico o contorno superior destacado entre o valor da
variável "teta" e o índice ambiental máximo
+ t<-t+1
+ text(teta,0,paste("T",t),cex=0.87, adj=c(0,1)) #atribui à
linha vertical "teta" o nome "Tt" (T2,3,...)
+   }
+ x[2]<-teta #o índice máximo passa a tomar o valor da
variável "teta"
+ for(j in 1:n)
+   {if (intercept [u]< intercept [j])
+     u<-j else if(intercept[u]== intercept [j] & va[u]+vb[u]*(
intercept [u])<va[j]+vb[j]*( intercept[j]))
+     u<-j else
+     u<-u
+   }
+ }
+ if (t<n) cat('\n Os restantes cultivares não fazem parte do
contorno superior!\n') # no final o programa irá verificar
que todos os cultivares compõem o Contorno Superior, caso não
se verifique envia a frase acima
+ legend("bottom",paste("Contorno Superior"),cex=0.7,lwd=2)
#coloca uma legenda a indicar o Contorno Superior
+ print('matriz com declive decrescente'); matriz_dec_b
+}

```


APÊNDICE B: TABELAS

Apêndice B: Tabelas de produção de cultivares (kg/hectare)

Tabela B. 1: Produção de cultivares (kg/ha) do ano de 2010

		1	2	3	4	5	6	7	8	9	10	11
		AI901	AI902	TS901	TS902	AI903	AI904	MVER	TS903	RCRUZ	TS904	BRAVO
VS	Rep1	667	1467	2000	2467	2333	2400	2067	2867	1867	1733	1467
	Rep2	1867	1667	2867	2533	1733	2333	1733	2267	2133	1800	1933
	Rep3	1667	1467	3533	3267	2600	2267	1800	2000	1533	1867	1533
	Rep4	1933	2133	3400	3733	2267	2267	2667	2067	2133	1533	1933
SA	Rep1	760	900	980	1313	780	880	633	993	647	553	513
	Rep2	867	860	1400	1333	1100	913	847	1000	773	820	940
	Rep3	267	1027	1960	1293	1047	820	1000	1393	867	1047	933
	Rep4	947	1160	660	1380	1080	833	627	1007	967	973	1113
SN	Rep1	2444	533	2000	1333	1867	667	1333	1933	1600	1200	1067
	Rep2	2000	2000	2667	1333	2667	2133	2533	2333	2000	1867	2000
	Rep3	2667	2133	3333	2667	2400	2000	2667	2933	2400	2667	2533
	Rep4	2667	2000	4000	3200	3467	2667	2933	2533	2667	2533	2533

Tabela B. 2: Produção de cultivares (kg/ha) do ano de 2011

		1	2	3	4	5	6	7	8	9	10	11
		AI901	AI902	TS901	TS902	AI904	AI905	MVER	TS903	RCRUZ	TS905	BRAVO
VS	Rep1	1466	1562	2417	1665	952	1071	1667	1190	1609	1786	1905
	Rep2	1400	1300	2210	1406	1190	1429	1667	1071	1629	1667	1667
	Rep3	996	888	1284	1269	1786	2024	714	1508	1226	2143	1905
	Rep4	1649	1461	1848	1710	1429	1429	1548	1667	1229	1786	2143
SA	Rep1	952	1190	1785	1428	1190	1508	1429	1905	1190	1607	2202
	Rep2	1071	1428	2023	1428	1190	1667	1310	2024	1507	1786	2024
	Rep3	1666	1666	714	1547	1190	1667	1548	1786	1428	1905	1786
	Rep4	1190	1071	1507	1666	1071	1071	1905	1548	1904	1786	1548
AMI	Rep1	595	1904	1547	1071	952	1667	2024	1429	833	2202	1905
	Rep2	1785	1666	2142	1785	1071	1190	1548	952	1607	1845	1825
	Rep3	1904	1666	1904	2142	714	833	1429	714	2202	1667	1786
	Rep4	1904	1904	2023	1190	583	280	292	658	1309	377	634
GOF	Rep1	738	642	523	535	643	298	889	952	428	274	266
	Rep2	928	857	642	928	214	351	986	457	821	1351	961
	Rep3	285	327	476	380	298	321	999	689	452	736	1496
	Rep4	583	642	214	297	298	643	1254	1245	297	736	1496

Tabela B.3: Produção de cultivares (kg/ha) do ano de 2012

		1	2	3	4	5	6	7	8	9	10	11
		AI901	AI902	TS901	TS902	AI904	AI905	ROSA	F.Fran ces	Sluz	TS905	BRAVO
VS	Rep1	1455	3337	3845	3012	3056	3792	3921	3304	3002	3606	1669
	Rep2	3984	3649	2884	2281	3900	3860	3790	4000	3902	3892	4513
	Rep3	4164	3942	3472	2394	3264	3309	3621	4221	4028	4289	5004
	Rep4	3432	3272	3873	3529	5422	3357	3656	4404	3840	3822	3773
VISTA	Rep1	3452	3273	2916	3869	4345	3154	3392	3511	3809	3511	3571
	Rep2	3452	3571	2380	3035	3690	2857	3095	4047	3511	3154	3214
	Rep3	3690	3690	3035	4285	5000	3511	3988	4107	3809	3750	4166
	Rep4	3333	2678	4404	4404	4642	3750	4285	4047	3452	3214	3928
AMI	Rep1	2713	3325	2424	1665	3042	2242	3509	3884	4544	4154	4412
	Rep2	2730	2561	2294	1984	4040	2830	3245	2986	3173	2742	3690
	Rep3	2614	3321	2693	2916	2580	3541	2108	2306	2038	2934	2578
	Rep4	3489	2060	3874	3816	4437	3941	3086	3798	2972	3001	3717
GOF	Rep1	976	1071	1071	166	750	369	714	357	333	785	476
	Rep2	1309	1309	714	1690	488.	702	1142	773	101	797	726
	Rep3	535	714	428	1380	1142	654	1166	833	952	928	500
	Rep4	976	1261.	833	1309	607	619	714	500	1142	952	1226

APÊNDICE C: MÉDIA DOS INDICES E R^2

Apêndice C: Média dos índices e R^2 dos anos de 2010 a 2012 nos respectivos locais.

Dados ano, local.

C.1: Resultados de 2010 por local

Dados ano 2010 no local VS

```
> cultivares<-read.table("cultivaresVS2010.txt",head=T)
```

Índices ambientais (x):

```
[1] 1939.545 2160.342 2275.853 2369.636
```

Média dos Índices ambientais (x):

```
[1] 2186.344
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 2045912 SS modelo: 2045912
```

```
[1] 2.664535e-15
```

Dados ano 2010 no local SA

```
> cultivares<-read.table("cultivaresSA2010.txt",head=T)
```

Índices ambientais (x):

```
[1] 851.4816 920.2564 1059.4545 813.8182
```

Média dos Índices ambientais (x):

```
[1] 911.2527
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 576480 SS modelo: 576480
```

```
[1] 2.953193e-14
```

Dados ano 2010 no local SN

```
> cultivares<-read.table("cultivaresSN2010.txt",head=T)
```

Índices ambientais (x):

```
[1] 1452.455 2160.223 2552.045 2836.364
```

Média dos Índices ambientais (x):

```
[1] 2250.271
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 2388019 SS modelo: 2388019
```

```
[1] 1.032507e-14
```

C.2: Resultados de 2011 por local

Dados ano 2011 no local VS

```
> cultivares<-read.table("cultivaresVS2011.txt",head=T)
```

Índices ambientais (x):

```
[1] 1627.182 1591.476 1431.182 1557.926
```

Média dos Índices ambientais (x):

```
[1] 1551.941
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 642153.7 SS modelo: 642153.7
```

```
[1] 1.121325e-14
```

Dados ano 2011 no local SA

```
> cultivares<-read.table("cultivaresSA2011.txt",head=T)
```

Índices ambientais (x):

```
[1] 1585.200 1587.091 1478.818 1526.481
```

Média dos Índices ambientais (x):

```
[1] 1544.398
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 1133432 SS modelo: 1133432
```

```
[1] 1.306732e-13
```

Dados ano 2011 no local AMI

```
> cultivares<-read.table("cultivaresAMI2011.txt",head=T)
```

Índices ambientais (x):

```
[1] 1583.273 1377.406 1299.624 1014.000
```

Média dos Índices ambientais (x):

```
[1] 1318.575
```

R2 diagnóstico (deve ser próximo de zero):

SS zigzag: 2821086 SS modelo: 2821086

[1] 1.454392e-14

Dados ano 2011 no local GOF

> cultivares<-read.table("cultivaresGOF2011.txt",head=T)

Índices ambientais (x):

[1] 562.5455 607.6244 748.4833 772.3636

Média dos Índices ambientais (x):

[1] 672.7542

R2 diagnóstico (deve ser próximo de zero):

SS zigzag: 1606548 SS modelo: 1606548

[1] 2.186029e-13

C.3: Resultados de 2012 por local**Dados ano 2012 no local VS**

> cultivares<-read.table("cultivaresVS2012.txt",head=T)

Índices ambientais (x):

[1] 3090.818 3779.474 3852.727 3644.512

Média dos Índices ambientais (x):

[1] 3591.883

R2 diagnóstico (deve ser próximo de zero):

SS zigzag: 5052587 SS modelo: 5052587

[1] 5.551115e-15

Dados ano 2012 no local VISTA

> cultivares<-read.table("cultivaresVISTA2012.txt",head=T)

APÊNDICE C

Índices ambientais (x):

```
[1] 3520.958 3273.273 3700.949 3911.909
```

Média dos Índices ambientais (x):

```
[1] 3601.772
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 1864300  SS modelo: 1864300
```

```
[1] 7.438494e-15
```

Dados ano 2012 no local AMI

```
> cultivares<-read.table("cultivaresAMI2012.txt",head=T)
```

Índices ambientais (x):

```
[1] 3471.909 3063.700 2693.545 2750.314
```

Média dos Índices ambientais (x):

```
[1] 2994.867
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 8549565  SS modelo: 8549565
```

```
[1] 1.354472e-14
```

Dados ano 2012 no local GOF

```
> cultivares<-read.table("cultivaresGOF2012.txt",head=T)
```

Índices ambientais (x):

```
[1] 642.5455 917.7683 921.7273 882.9139
```

Média dos Índices ambientais (x):

```
[1] 841.2387
```

R2 diagnóstico (deve ser próximo de zero):

```
SS zigzag: 2014911  SS modelo: 2014911
```

```
[1] 5.306866e-14
```