

UNIVERSIDADE ABERTA
Departamento de Ciências e Tecnologia

Análise Conjunta de Regressões
com desenvolvimentos computacionais em Linguagem R

Aluno: Márcio do Carmo Martins

Dissertação apresentada à Universidade Aberta para obtenção do grau de Mestre em Estatística, Matemática e Computação, especialidade Estatística Computacional

Orientador: Professor Doutor Amílcar Manuel do Rosário Oliveira
(Professor Auxiliar da Universidade Aberta)

Lisboa

2010

Agradecimentos

Ao professor doutor Amílcar Manuel do Rosário Oliveira, meu orientador, agradeço pelo apoio, pela paciência e pelo constante incentivo na realização desta dissertação permitindo o contacto com uma nova área da estatística computacional.

Aos professores da parte curricular do mestrado agradeço pela disponibilidade e pelo constante incentivo na realização de todos os trabalhos que nos foram propostos.

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

A toda a minha família o apoio incondicional e compreensão que tiveram pela minha ausência em diversos momentos.

À Prof^a. Sandra Cruz agradeço o seu apoio em todos os momentos.

Aos meus colegas de trabalho agradeço a paciência e apoio em momentos mais complicados. Agradeço sobretudo pelo trabalho adicional em alguns momentos que tive de me ausentar.

Resumo

R é uma linguagem e um ambiente de desenvolvimento integrado, para cálculos estatísticos e gráficos. R é também altamente expansível com o uso dos pacotes, que são bibliotecas para funções específicas ou áreas de estudo específicas.

Por outro lado, a Análise Conjunta de Regressões (ACR) é uma técnica com bastante aplicabilidade na comparação e selecção de cultivares. Na ACR existe um conjunto de procedimentos que permitem fazer a comparação e selecção de cultivares.

A técnica ACR consiste no ajustamento de regressões lineares do rendimento de cada cultivar no respectivo índice ambiental. Para o mesmo ajustamento, usualmente recorremos ao algoritmo Zig-Zag.

Após a obtenção dos coeficientes ajustados, podemos através do Método do Contorno Superior, mostrar quais os cultivares que integram esse contorno, num determinado intervalo de índices ambientais.

Neste trabalho, após o enquadramento da técnica ACR, baseada em Oliveira (2007) e Pinto (2006), pretende-se o desenvolvimento de procedimentos e de um programa em linguagem R que permita a obtenção do Contorno Superior, indicação gráfica desse Contorno e dos cultivares dominantes (que integram esse contorno superior) num determinado intervalo de índices ambientais.

Por fim, aplicaremos o mesmo programa a um conjunto de dados reais para verificação da aplicabilidade do mesmo.

Abstract

R is a language and an integrated development environment for statistical calculations and graphs. R is also highly expandable with the use of packages, which are libraries for specific functions or specific areas of study.

Moreover, the Joint Regression Analysis (JRA) is a technique with great applicability for the comparison and selection of cultivars. In JRA there is a set of procedures that allow the comparison and selection of cultivars.

The JRA technique consists in the adjustment of linear regressions of the yield of each cultivar in their environmental index. For the same adjustment, usually resort to the Zig-Zag algorithm.

After obtaining the adjusted coefficients, we can through the Upper Contour Method, which show the cultivars belonging to that Contour, in a certain range of environmental indexes.

In this work, after to frame the JRA, based in Oliveira(2007) and Pinto(2006), with intend to develop procedures and a program in R language that allows obtaining the Upper Contour, an graphic indication of that Contour and dominant cultivars (incorporating the Upper Contour) at a range of environmental indexes.

Finally, we apply the same program to a real data set to verify the applicability of the same.

Lista de Símbolos e Siglas

ACR	Análise Conjunta de Regressões
β	Coefficiente angular de regressão linear
L_2	Índices ambientais L_2
J	Número de cultivares estudados
b	Número de blocos
$p_{ij}; i = 1, \dots, b; j = 1, \dots, J$	Peso atribuído ao j – <i>ésimo</i> cultivar no i – <i>ésimo</i> bloco
$p_i; i = 1, \dots, b$	Peso atribuído a um cultivar que esteja presente no bloco i
$y_{ij}; i = 1, \dots, b; j = 1, \dots, J$	Produção do j – <i>ésimo</i> cultivar no i – <i>ésimo</i> bloco
α_j	Ordenada na origem da recta de regressão a ajustar do j – <i>ésimo</i> cultivar
β_j	Declive da recta de regressão a ajustar do j – <i>ésimo</i> cultivar
x_i	Índice ambiental do i – <i>ésimo</i> bloco
α_J	Vector das ordenadas na origem para as regressões
β_J	Vector dos declives para as regressões
x_b	Vector dos índices ambientais
$S_1(\alpha_J, \beta_J)$	Função objectivo a minimizar na primeira iteração do algoritmo Zig-Zag
$\mathbb{R}_b \times \mathbb{R}_J \times \mathbb{R}_J$	Espaço obtido pelo produto cartesiano
x_i^0	Valor inicial do índice ambiental do i – <i>ésimo</i> bloco

$R(\mathbf{1}_b)$	Espaço imagem dos vectores coluna com todas as componentes iguais a 1
$grad(S)$	Gradiente da função S
$Hess(S)$	Matriz hessiana de S
$(\tilde{\alpha}_j, \tilde{\beta}_j)$	Coefficientes de regressão ajustados referentes ao j – ésimo cultivar
\tilde{x}_i	Índice ambiental ajustado no i – ésimo bloco
$r_{i,j}$	Resíduo associado ao j –ésimo cultivar no i – ésimo bloco
L	Número de sub-redes conectadas
$A_l; l = 1, \dots, L$	Conjunto dos índices dos blocos para a l – ésima sub-rede conectada.
$(\tilde{\alpha}_j^1, \tilde{\beta}_j^1)$	Estimadores dos mínimos quadrados para α_j^1, β_j^1 obtidos na primeira iteração do algoritmo Zig-Zag
$S_1(\mathbf{x}_b)$	Função a minimizar em ordem a \mathbf{x}_b na primeira iteração do algoritmo Zig-Zag
$Int(a)$	Parte inteira de a
$A \otimes B$	Produto de Kronecker de A por B
\mathbf{x}'_b	Transposta de \mathbf{x}_b
\mathbf{Y}'	Transposta de \mathbf{Y}
$L(.)$	Função de máxima verosimilhança
$D_i; i = 1, \dots, b$	Conjunto dos índices dos cultivares presentes no i – ésimo bloco
N	Número de observações

$\ln(.)$	Logaritmo da função de verosimilhança
$\hat{\sigma}^2$	Estimador de máxima verosimilhança para σ^2
$(\hat{\alpha}_j, \hat{\beta}_j)$	Estimadores de máxima verosimilhança
$\mathbf{1}_J$	Vector com J componentes iguais a 1
$\tilde{\theta}_{j,j'}$	Ponto de intersecção entre as rectas de regressão ajustadas com índices j e j'
$H_0^{j,j'}(\theta_0)$	Hipótese nula relativa aos j – ésimo e j' – ésimo cultivares e à função $U(\theta_0)$
$N(\underline{\eta}_m, \sigma^2 K)$	Distribuição normal, com valor médio $\underline{\eta}_m$ e matriz de covariância $\sigma^2 K$
χ_g^2	Distribuição qui-quadrado central, com g graus de liberdade
$D(r_1, \dots, r_m)$	Matriz diagonal do tipo $s \times s$, cujos elementos da diagonal principal correspondem a r_1, \dots, r_m
S_{a_l}	Soma dos quadrados dos desvios para a média dos índices ambientais ajustados no ano com índice a_l

Conteúdo

Capítulo 1	1
Introdução	1
Capítulo 2	3
Linguagem de Programação R	3
2.1. Nota histórica	3
2.2. Ambiente de Programação R	7
Capítulo 3	15
Análise Conjunta de Regressões (ACR)	15
3.1. Nota histórica sobre a ACR.....	15
3.2. Caso geral	17
3.2.1. Função objectivo	17
3.2.2. Redes de ensaios conectadas	20
3.2.3. Algoritmo Zig-Zag	22
3.3. Caso completo.....	25
3.3.1. Técnica de Dupla Minimização	26
3.3.2. Estimadores de Máxima Verosimilhança	29
3.3.3. Testes de razão de verosimilhanças.....	30
3.4. A importância da Estabilidade dos Cultivares.....	35
Capítulo 4.....	38
Método do Contorno Superior	38
4.1. Selecção pelo método do Contorno Superior.....	38
4.1.1. Geometria.....	39
4.2. ACR Lineares Completadas	54
Capítulo 5.....	55
Método prático para detecção do Contorno Superior	55
5.1. Algoritmo.....	55
5.2. Fluxograma.....	58
5.4. Exemplo de Aplicação	59
Conclusões e trabalho futuro	70

Anexo	71
Anexo A.....	71
Programa em R	71
Bibliografia.....	78

Lista de Figuras

Figura 4.1. Traçado do contorno superior para duas rectas com declives diferentes. ...	41
Figura 4.2. Traçado do contorno superior para duas rectas com declives iguais.	42
Figura 4.3. Traçado do contorno superior para quando $\bar{\theta}_1 \leq \theta_{min}$	44
Figura 4.4. Traçado do contorno superior para quando $\theta_{min} < \bar{\theta}_1 < \theta_{max}$	46
Figura 4.5. Traçado do contorno superior para quando $\theta_{max} \leq \bar{\theta}_1$	48
Figura 4.6. Intersecção de várias rectas com declives diferentes.	50
Figura 4.7. Intersecção de três rectas com declives diferentes no mesmo ponto.	53
Figura 5.1. Aplicação do Método do Contorno Superior aos dados de 1993 (J=11).	61
Figura 5.2. Aplicação do Método do Contorno Superior aos dados de 1994 (J=11).	63
Figura 5.3. Aplicação do Método do Contorno Superior aos dados de 1995 (J=11).	65
Figura 5.4. Aplicação do Método do Contorno Superior aos dados de 1996 (J=11).	67
Figura 5.5. Aplicação do Método do Contorno Superior aos dados de 1997 (J=11).	69

Capítulo 1

Introdução

A Estatística é o ramo da Matemática que coleciona, classifica, analisa e interpreta dados quantitativos ou qualitativos com o intuito de tomar decisões.

Neste seguimento podemos encontrar 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 falta de conhecimento da existência da mesma ou o facto de ser bastante exaustiva e demorada leva muitos investigadores a optarem por outras alternativas.

De entre as diversas técnicas de análise de dados existe a Análise Conjunta de Regressões, ACR. De acordo com Oliveira, A. (2008) esta técnica tem sido utilizada na análise conjunta de ensaios de comparação de cultivares.

"A técnica ACR, consiste no ajustamento de regressões lineares do rendimento de cada cultivar no respectivo índice ambiental" (Pinto, 2006).

A Análise Conjunta de Regressões, ACR, pertence ao grupo das técnicas baseadas no ajustamento de regressões.

Dentro da Análise Conjunta de Regressões, ACR, Mexia et al. (1997) introduziu a utilização do método do contorno superior no processo de comparação e selecção de cultivares. Neste método devemos seleccionar o contorno superior resultante da intersecção das rectas de regressão linear associadas aos cultivares em estudo.

Consequentemente, podemos verificar que o contorno superior é uma poligonal convexa pois a mesma é constituída por segmentos pertencentes às regressões dominantes. Quando duas rectas se intersectam num determinado ponto (ponto de intersecção), à direita do mesmo ponto, a recta dominante é a que possui maior declive e à esquerda, do mesmo, a recta dominante é a que possui menor declive. Por conseguinte, verifica-se que os coeficientes angulares dos troços que constituem o contorno superior, crescem da esquerda para a direita formando a poligonal convexa.

Deste modo, *"a informação fornecida pelo traçado do contorno superior, permite facilmente identificar um conjunto de cultivares, que se destacam do grupo pelo facto de apresentarem produções máximas, para certos valores do índice ambiental"* (Pinto, 2006).

Perante este facto, pretendemos criar um método expedito que permita automatizar alguns processos, nomeadamente no que diz respeito ao método do contorno superior. Essa automatização permitir-nos-á obter de forma célere e eficiente a informação fornecida pelo traçado do contorno superior, tendo como base os dados referentes aos cultivares que se pretendam analisar. Para a elaboração do método expedito iremos recorrer ao software *OpenSource R*, bastante versátil, que será descrito de forma sucinta no Capítulo 2. Daremos ênfase especial às funções que serão utilizadas nos Capítulos 4 e 5.

No Capítulo 3 será feita uma retrospectiva sobre a Análise Conjunta de Regressões e focalizados as apectos mais importantes desta técnica.

No capítulo 4 será dado especial destaque à apresentação do Método do Contorno Superior.

Finalmente no Capítulo 5 iremos desenvolver o algoritmo para o Método do Contorno Superior e aplicação do mesmo a dados reais.

Capítulo 2

Linguagem de Programação R

2.1. Nota histórica

A linguagem de programação R é um ambiente de linguagem e de *software* de programação para cálculos estatísticos e gráficos. 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.

Outra vantagem que a mesma linguagem de programação dispõe é a sua disponibilidade de forma gratuita a todos os interessados e permite que os mesmos desenvolvam novos modelos estatísticos e que os mesmos sejam disponibilizados aos restantes investigadores ou curiosos. Por conseguinte, podemos afirmar que se trata de uma linguagem em constante evolução.

A linguagem de programação R apareceu inicialmente em 1996, quando os professores de Estatística Ross Ihaka e Robert Gentleman, da Universidade de Auckland, Nova Zelândia, começaram a distribuir o código como um software de fonte aberta, daí a designação R pois são as iniciais dos primeiros nomes dos seus criadores e foi desenvolvido por um esforço colaborativo de pessoas em vários locais do mundo.

No entanto, esta linguagem começou a ser desenvolvida no início dos anos 90, mais propriamente em 1991. Os mesmos professores demoraram cerca de 5 a 6 anos a desenvolver a mesma linguagem pois não tinham muita experiência na computação.

"Por cinco ou seis anos, trabalhávamos juntos o tempo todo... Um digitava e o outro pensava", referiu Gentleman ao jornal New York Times em 2009 (ver www.nytimes.com/2009/01/07/technology/business-computing/07program.html).

A primeira distribuição do R foi disponibilizada em 1993, quando algumas cópias foram disponibilizadas no StatLib, um sistema de distribuição de *softwares* estatísticos.

A linguagem R tem segundo vários autores, como base a linguagem de programação estatística S, no entanto pode ser considerada uma implementação distinta do S;

embora com importantes diferenças, muitos códigos escritos para o S podem ser usados inalterados no R.

Os dois professores desejavam tecnologia mais adaptada às necessidades dos seus alunos de estatística, que precisam de analisar dados e produzir modelos gráficos dessas informações. Outro facto que os levou a optarem por criarem a sua própria linguagem estava ligado com a complexidade dos softwares estatísticos que estavam disponíveis, sendo que os mesmos tinham sido criados por cientistas da computação que não tinham tanta sensibilidade para a vasta área que é a estatística.

De acordo com uma notícia publicada no New York Times (6 de Janeiro de 2009), *"para algumas pessoas, o R é apenas a 18ª letra o alfabeto"*. Contudo a linguagem está a tornar-se numa linguagem padrão devido aos seus grandes potenciais e o seu uso está em crescente e está a ser usada, cada vez mais, por um número significativo de analistas de dados, em empresas e no mundo académico.

Outro facto que leva a linguagem R estar em constante evolução está relacionado, segundo o mesmo artigo, com a constante recolha e tratamento de dados, motivada por meios como, a publicidade, a descoberta de novos medicamentos de forma mais rápida, modelos financeiros entre outros.

O New York Times também salientou que a linguagem R *"encontrou rápida aceitação entre os estatísticos, engenheiros e cientistas que não conhecem bem a programação de computadores"* pois os mesmos consideraram a linguagem bastante fácil de utilizar. *"A aplicação R é realmente importante, tanto que é difícil superestimá-la... Ela permite que os estatísticos realizem análises muito intrincadas e complicadas sem que precisem conhecer em detalhe o funcionamento dos sistemas de computação"* referiu Daryl Pergibon, cientista da Google, na mesma notícia.

Por outro lado, podemos também considerar que a linguagem R é semelhante a outras linguagens de programação, como C, Java e Perl, porque ajuda as pessoas a executar ampla variedade de tarefas de computação, sendo que, este facto está relacionado com o fornecimento de diversos comandos por parte das mesmas. Para os estatísticos, a aplicação 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.

A informação disponibilizada em vários motores de busca na internet é bastante vasta, cerca de dois mil pacotes (*packages*) podem se encontrar em apenas um dos muitos sites dedicados à linguagem R, número que tem vindo a crescer exponencialmente.

"A grande beleza da aplicação R é que se pode modificá-la para realizar diversas coisas diferentes", disse Hal Varian, economista chefe do Google, ao New York Times (2009).

No âmbito da mesma notícia publicada pelo New York Times, 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"*. Este facto revela outra vantagem da linguagem R que está ligada ao ser *Open Source*.

O código fonte do R está disponível sob a licença GNU GPL e as versões binárias pré-compiladas são fornecidas para Windows, Macintosh, e para outros sistemas operacionais como o Unix/Linux.

Motivado pela constante evolução da linguagem R, a mesma é também altamente expansível com o uso dos *packages*, que são bibliotecas para funções específicas ou áreas de estudo específicas, por conseguinte dependendo das necessidades do utilizador podemos acrescentar funcionalidades ao programa, caso essas mesmas funcionalidades não existam, podem ser criadas pelo próprio utilizador de uma forma mais simples ou complicada consoante o que o mesmo pretender.

Devido aos potenciais acima indicados, grandes empresas estão a render-se ao R. Um dos exemplos desse mesmo facto é a Google que está a usar esta linguagem em muitos estudos. E mais recentemente uma outra grande multinacional também se rendeu à mesma, a Oracle vai começar a dar suporte à linguagem de código aberto R, para análise estatística.

Segundo a ComputerWorld, *"já está disponível uma interface de utilizador para o software de Data Mining da Oracle, dirigido a profissionais de estatísticas e de análise de dados que trabalham com R mas que não têm tão bons conhecimentos de SQL ou da tecnologia de base de dados de Oracle... A nova interface vai, assim, juntar-se à*

comunidade existente de desenvolvimento de packages R para Oracle, incluindo um driver da base de dados e uma interface de utilizador. A nova interface da Oracle é totalmente externa ao seu software, consistindo em "cápsulas" de funções que transferem informações entre o ambiente R e a base de dados na forma de consultas PL/SQL". Contudo "embora a interface possa ser descarregada gratuitamente, os utilizadores da linguagem R terão sempre que investir em produtos da Oracle".

Também de acordo com a revista ComputerWorld, *"reconhecendo igualmente a crescente influência e potencial de adopção da linguagem de código aberto R, também outras empresas concorrentes da Oracle, como a IBM ou a SAS, e inclusive companhias de menor dimensão, como a Revolution Analytics, decidiram oferecer software e serviços para esta linguagem".*

Segundo Torgo (2009), uma grande parte das empresas produz um volume de dados de dimensão muito elevada e com tendência crescente como resultado do registo das suas actividades. Neste seguimento, a análise destes dados é de primordial importância para otimizar a tomada de decisões nas mesmas. Deste modo pretendemos criar um novo programa que permite aplicar o método do contorno superior de forma simplificada e organizada.

É de salientar que este programa irá enriquecer a linguagem R, somando-lhe uma nova funcionalidade e aplicabilidade prática no estudo do melhoramento de cultivares. Outra vantagem é a flexibilidade na gestão de grandes quantidades de dados.

No entanto ao compararmos este software a outros, o R tem a desvantagem de não possuir uma interface gráfica suficientemente elaborada que permita uma interacção amigável com o utilizador, o que pode, de início, desestimular a sua utilização, já que a sua aprendizagem pode tornar-se um pouco mais lenta que com outros programas.

Em R devemos digitar na linha de comando as funções necessárias à realização das análises. Deste modo confere-lhe a uma grande flexibilidade tal como referimos acima, no entanto a aprendizagem é mais difícil.

2.2. Ambiente de Programação R

Como referimos anteriormente, a linguagem de programação R está disponível a qualquer utilizador que disponha de ligação à internet. Em qualquer motor de busca na internet ele pode ser encontrado para download, sendo que, podemos ir directamente ao site “www.r-project.org” e encontrar aí para além do programa, muita informação sobre o mesmo. Mas não é obrigatório que um utilizador tenha sempre acesso à internet quando está a usar o R.

O utilizador pode também aceder à ajuda disponível no mesmo programa, basta escrever `help()`. Através deste comando irá aparecer todas as informações que o programa dispõe, pois entre parênteses não colocamos nenhum argumento ou fazer uma pesquisa mais objectiva, como por exemplo encontrar o comando para realizar a raiz quadrada “`help(sqrt)`”.

No entanto visto que é um projecto em crescimento, muitas opções podem não se encontrar disponíveis e neste caso temos de recorrer à internet e fazer uma pesquisa sobre *packages* do software R. Como referimos acima, milhares de *packages* estão disponíveis em diversos sites. Apenas necessitamos de realizar o download dos mesmos e depois instalá-los dentro do software R através do comando “`install.packages()`” colocando no argumento o nome do *package* a instalar, ou em alternativa podemos ir ao menu **Packages** e seguidamente ir à opção “**install package(s)**” onde surgirão diversos *packages* disponíveis.

O mesmo *package* fica automaticamente disponível se recorrermos à opção “**install package(s)**” do menu **Packages**. Seguidamente devemos escrever o comando `library()` colocando novamente o nome do *package* no argumento para nos certificarmos que o *package* está devidamente instalado. Se somente usarmos `library()` sem argumento apenas nos é listado todos os *packages* que estão em funcionamento.

No presente trabalho irá ser necessário a instalação do *package* “`gdata`” que permite a importação de dados de tabelas de Excel, Sql e mesmo de dados no “clipboard” do

computador que estamos a trabalhar de um "Copy" dos dados de uma tabela de Excel e seguidamente através da seguinte instrução no R:

```
> dados<-read.table('clipboard', header=T)
```

onde o R irá importar os valores que foram copiados da tabela e que posteriormente podemos analisar, sendo que a primeira linha do arquivo tem as descrições das colunas que são adicionamos pelo *header=T* e o R vai usar o conteúdo desta mesma linha para atribuir nomes às colunas.

Outra instrução que podemos realizar aquando da introdução de dados é a seguinte

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

onde o ficheiro com os dados encontra-se em formato *.txt* e o R irá introduzir esses dados considerando a primeira linha de dados como o título (*header=T*). A importação de dados pode ser efectuada de outras bases de dados, como SQL, SPSS, devemos nesse caso indicar no lugar de "c:\\dados.txt", a indicação da base de dados e respectiva extensão.

Outro elemento que nos pode ser bastante útil no tratamento de dados é a sua ordenação porque os mesmos podem não estar ordenados na tabela de origem. Podemos ordenar os dados de ordem ascendente, como descendente recorrendo a uma instrução como está em seguida

```
>dados_ord<-dados[order(dados$x),]
```

Esta instrução ordena os dados numa nova lista de forma ascendente segundo os valores "x" da tabela dados. Contudo poderíamos estar interessados numa ordem descendente e nesse caso deveríamos recorrer à instrução

```
>dados_ord<-dados[order(-dados$x),]
```

onde o "-" apenas indica o sentido da ordenação dos dados. Por outro lado, poderíamos usar a instrução "sort"

```
>dados_ord<-sort(dados$x,decreasing=FALSE)
```

onde a ordenação dos dados fica de forma ascendente

```
>dados_ord<-sort(dados$x,decreasing=TRUE)
```

onde a ordenação dos dados fica de forma descendente.

Podemos usar outras instruções que permitem a indicação do valor máximo, ou mínimo, de uma lista de valores, sendo um exemplo a instrução seguinte onde iria ser indicado o máximo dos valores de "x" na tabela "dados"

```
> max(dados$x)
```

ou a instrução seguinte que indicaria o mínimo da mesma lista de valores

```
> min(dados$x)
```

Por outro lado, se pretendemos que o programa nos devolva uma frase devemos usar a seguinte instrução

```
> print('frase que se pretende...')
```

onde entre aspas devemos escrever o que o programa deve devolver. O programa neste caso devolve uma frase tal como foi solicitado pelo programador.

No entanto se pretendemos que o programa nos devolva uma frase em que alguns dos argumentos são dados do próprio programa devemos usar a instrução `cat()` que funciona com um número qualquer de argumentos, e o que faz é transformar os seus argumentos em "strings", concatená-los, e só depois os escreve no ecrã. De acordo com Torgo (2009) apresentamos um exemplo da mesma instrução

```
> x<-34
```

```
> cat('x tem o valor de',x,'o que é estranho!')
```

```
x tem o valor de 34 o que é estranho!
```

Se utilizarmos a instrução `print(.)` o programa efectua uma quebra de linha após a devolução da frase, com recurso à instrução `cat(.)` o mesmo não se verifica devolvendo todas as respostas de forma continua. Quando pretendemos uma quebra de linha aquando do recurso à instrução `cat(.)`, devemos

```
> x<-34  
  
> cat('\n x tem o valor de',x,'o que é estranho!')  
  
x tem o valor de 34 o que é estranho!
```

A inclusão de `\n` seguido à aspa provoca a quebra de linha.

Mas, tal como referimos acima, existem muitas mais opções disponíveis neste programa tal como as análises de regressão.

Possuindo uma base de dados num determinado programa e, tal como referimos acima, podemos importar os dados para o R, seguidamente podemos fazer uma nuvem de pontos com os valores tabelas recorrendo à instrução

```
> plot(dados$x,dados$y)
```

e podemos calcular o coeficiente de correlação linear de Pearson entre as variáveis

```
> cor(dados$x,dados$y)
```

Queremos também ajustar um modelo que relacione as variáveis x e y da base de dados, sendo que no presente caso estamos interessados numa regressão linear simples do tipo $y = a*x + b$. A função `lm()` é a indicada para o mesmo objectivo

```
> lm(dados$y~dados$x)
```

Note com a função `lm()` é chamada com o formato `lm(y ~ x)` - lemos isto como *y depende de x*. Esta é a forma básica na qual diversos modelos são construídos, com a variável que está sendo medida (resposta) como a valor de y e as quantidades controladas (preditores) como as variáveis x .

Assim como a maioria das funções do R, podemos armazenar os resultados devolvidos pela função `lm()` num objecto. O valor devolvido pela instrução `lm()` é uma lista:

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

```
[1] "coefficients" "residuals" "effects" "rank"
[5] "fitted.values" "assign" "qr" "df.residual"
[9] "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`. Podemos usá-lo para adicionar uma linha ao gráfico de pontos:

```
> plot(dados$x,dados$y)
> valores$coef
> abline(valores$coef)
```

A função `abline()` recebe um vector com dois elementos e adiciona uma recta ao gráfico com estes valores de ordenada e inclinação, respectivamente. Se pretendemos apenas uma recta vertical no gráfico devemos indicar

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

se pretendemos uma recta horizontal devemos substituir o `v` por `h` na instrução acima.

Outros componentes do ajuste são úteis para o diagnóstico. Por exemplo pode-se examinar o gráfico de valores ajustados contra resíduos:

```
> plot(valores$fitted,valores$resid)
```

que pode mostrar os desvios dos dados em relação ao modelo linear, caso hajam curvaturas claras no gráfico.

Podemos também fazer um histograma dos resíduos para verificar a presença de outliers.

```
> hist(valores$resid)
```

O R produz automaticamente uma série de gráficos para o ajuste dum modelo linear. Ao introduzirmos o comando abaixo serão mostrados diversos gráficos relacionados ao ajuste do modelo linear:

```
> plot(sfit)
```

Outras instruções de bastante aplicabilidade e que iremos usar na concepção do programa da presente dissertação são as estruturas de controlo. Segundo Torgo (2009), a linguagem de programação R, tal como qualquer linguagem de programação, possui várias instruções destinadas a alterar o curso sequencial normal de execução dos programas.

Na presente dissertação iremos usar dois tipos de instruções de controlo, as instruções condicionais que permitem explicitar diferentes alternativas a serem executadas dependendo de alguma condição a ser testada na altura da execução das instruções e as instruções iterativas que nos permitem repetir blocos de instruções, sendo que neste caso o R possui ainda duas instruções que podem ser usadas para controlar a execução dos ciclos.

Dentro das instruções condicionais, segundo Torgo (2009) a instrução *if* permite explicitar uma condição booleana e dois conjuntos de instruções alternativos que são executados dependendo do valor da condição. A sua sintaxe genérica é

```
if <condição> <instrução> else  
  <instrução>
```

Nas instruções iterativas as instruções mais usadas são o "*for*", o "*while*" e o "*repeat*".

A instrução "*for*" permite controlar o número de vezes que um ciclo é executado através de uma variável de controlo que vai percorrer todos os valores de uma série pré-definida de valores, em cada iteração do ciclo, sendo que a sua sintaxe genérica é

```
for (<var> in <conjunto>)  
    <bloco de instruções>
```

A instrução "while" tem a sintaxe genérica

```
while (<condição booleana>)  
    <bloco de instruções>
```

De acordo com Torgo (2009), a sua semântica pode ser descrita com enquanto a condição booleana for verdadeira, repetir o bloco de instruções sendo que as mesmas podem nunca ser executadas, bastando para isso que a condição booleana seja falsa aquando da entrada no ciclo.

A instrução "repeat" permite mandar o programa R executar um bloco de instruções uma ou mais vezes, sendo a sua sintaxe genérica

```
repeat  
    <bloco de instruções>
```

De acordo com Torgo (2009), podemos verificar que o ciclo "repeat" não possui nenhuma condição lógica a governar a execução repetida do bloco de instruções, sendo que a instrução *break* permite que o mesmo ciclo pare assim que uma condição se verifique.

Recorrendo à instrução *stop* o programa pára independentemente da existência de mais ciclos dentro do corpo de mesmo.

Além das instruções referidas acima podemos usar outras de acordo com as necessidades do programa que estamos a desenvolver. Por outro lado podemos criar instruções condicionais dentro de instruções condicionais ou de instruções iterativas e vice-versa. Segundo Torgo (2009)

```
> x<-rnorm(10)
> k<-0
> for (v in x){
+     if (v > 0)
+         k <- k + v}
```

é um exemplo onde podemos usar uma instrução condicional (*if*) dentro de uma instrução iterativa (*for*).

Algumas instruções que descrevemos acima (ciclo condicional *if* e os ciclos iterativo *for* e *while*) serão usadas no programa que iremos desenvolver para indicar o contorno superior que indicaram os cultivares dominantes num experimento.

Todo o programa irá ficar inserido numa função (*function*). A criação da mesma consiste na atribuição do conteúdo de uma função a um nome, sendo que esse conteúdo é a lista dos seus argumentos e as instruções que formam o corpo da função.

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

```
<nome da função> <- function(<lista de argumentos>)
    <corpo da função>
```

Por conseguinte a função irá devolver o resultados das instruções que se encontram no corpo da mesma.

Muitas outras instruções e funções, assim como *packages* estão disponíveis contudo, tal como em todos os programas, devemos praticar bastante para que tenhamos maior agilidade no uso do mesmo. Por conseguinte reforçamos que deve-se praticar bastante, pois é um programa bastante versátil e de grande aplicabilidade.

Capítulo 3

Análise Conjunta de Regressões (ACR)

3.1. Nota histórica sobre a ACR

A análise de regressões é uma técnica conhecida pela sua utilização na comparação e selecção de cultivares.

Segundo Oliveira (2007), a técnica tem sido utilizada na interpretação de redes de ensaios destinados à comparação de cultivares. O mesmo autor refere que a técnica foi originalmente descrita por Mooers (1921), tendo sido introduzida por Yates & Cochran (1938), onde calcularam a média total de todos os cultivares em determinado ambiente e designaram essa medida por índice ambiental. Deste modo, Yates & Cochran estudavam o comportamento de um determinado cultivar, com recurso a uma regressão linear das produções do mesmo nos índices ambientais.

Posteriormente, no decorrer da década de sessenta a mesma técnica ACR sofreu grandes avanços científicos através de muita investigação de diversos investigadores, sendo que Finlay & Wilkinson (1963), Eberhart & Russel (1966) e Perkins & Jinks (1968) foram alguns dos grandes investigadores da década. Segundo Pinto (2006), os mesmos investigadores deram o seu grande contributo para aplicação da técnica à comparação de cultivares sujeitos a um vasto número de níveis de fertilidade.

Contudo, de acordo com Oliveira (2007), apesar das constantes evoluções da técnica existem algumas limitações da mesma. Na década de 80, Wescott (1986) e Lin *et al.* (1986) indicaram que a ACR não tinha em consideração 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 dezena e meia.

Também no trabalho de Romagosa & Fox (1993) verifica-se que a técnica ACR revela fragilidade quando os ambientes testados são restritos em número e pouco diversos ou quando os cultivares incluídos na experiência já foram bastante seleccionados impedindo divergências de valores.

De acordo com Oliveira (2007), em Portugal a investigação na área da ACR começou a ter evidencia na década de oitenta com trabalhos de Gusmão (1985), (1986a) e (1986b). Entre outras conclusões este autor demonstrou que a consideração de distintos índices ambientais para os diferentes blocos, em detrimento de apenas um índice ambiental por ensaio, conduz a uma melhoria da precisão dos ajustamentos. Outros autores também fizeram trabalhos relevantes para a evolução da técnica no nosso país.

Oliveira (2007) refere que os trabalhos inicialmente apresentados por Gusmão (1985) e (1986a) reflectiam aplicações da ACR a redes de ensaios em blocos casualizados, em que todos os índices ambientais correspondiam aos rendimentos médios dos diversos blocos. Por conseguinte, para cada bloco era calculada a média das produções ajustando-se, cultivar a cultivar, regressões lineares das produções médias. Mas esta técnica tinha uma limitação provocada pela sua aplicabilidade apenas ser possível em situações de blocos completos casualizados e também surgem algumas objecções quanto à consideração das produções médias como valores de uma variável controlada.

Mas as mesmas limitações foram ultrapassadas no trabalho de Mexia et al. (1999) com a introdução dos índices ambientais L_2 e extensão das técnicas da ACR aos planos em blocos incompletos. Na dissertação de doutoramento de Pereira (2004) desenvolve-se de forma sistemática a ACR utilizando os índices ambientais L_2 sendo que a mesma aplicação é efectuada com recurso a um algoritmo e permite obter estimadores de máxima verosimilhança dos índices ambientais e a obtenção destes estimadores para outros parâmetros e a realização de testes de razão de verosimilhança. Por conseguinte consegue-se a integração da ACR na inferência estatística.

3.2. Caso geral

Ao realizarmos o ajustamento com recurso à ACR devemos ter uma função objectivo que pretendemos minimizar, sendo que os argumentos da mesma função são os índices ambientais e os coeficientes resultantes das regressões lineares simples.

Por conseguinte, inicialmente iremos apresentar a função objectivo e seguidamente o algoritmo Zig-Zag, sendo que o mesmo algoritmo é iterativo e permite no final de cada iteração transformar os índices ambientais ajustados para manter invariante o seu intervalo de variação.

3.2.1. Função objectivo

Considerando $p_{i,j} = 0$ se no i -ésimo bloco não estiver presente o j -ésimo cultivar e $p_{i,j} = 1$ se no i -ésimo bloco estiver presente o j -ésimo cultivar. Pretendemos minimizar

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2 \quad (3.2.1)$$

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

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

A escolha dos valores iniciais para as componentes do vector \mathbf{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 $x_i^0, i = 1, \dots, b$, os valores iniciais para os índices ambientais, iniciamos com a minimização de

$$S(\boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i^0)^2 \quad (3.2.2)$$

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

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

Lema 3.1. Sendo $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J)$ quando existirem ε e $\delta \neq 0$ tal que

$$\begin{cases} \mathbf{x}'_b = \delta \mathbf{x}_b + \varepsilon \mathbf{1}_b \\ \boldsymbol{\alpha}'_J = \boldsymbol{\alpha}_J - \frac{\varepsilon}{\delta} \boldsymbol{\beta}_J \\ \boldsymbol{\beta}'_J = \frac{1}{\delta} \boldsymbol{\beta}_J \end{cases} \quad (3.2.3)$$

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

Dem. Considerando o caso em que $\delta = 1$ e $\varepsilon = 0$ teremos para todo o trio $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ e conseqüentemente τ goza da propriedade reflexiva.

Admitindo que $(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) \tau (\mathbf{x}'_b, \boldsymbol{\alpha}'_J, \boldsymbol{\beta}'_J)$ e existindo $\delta \neq 0$ e ε tais que (3.2.3) se verifica, então sendo $\delta' = \delta^{-1}$ e $\varepsilon' = -\frac{\varepsilon}{\delta}$ tem-se

$$\begin{cases} \mathbf{x}_b = \delta' \mathbf{x}'_b + \varepsilon' \mathbf{1}_b \\ \boldsymbol{\alpha}_J = \boldsymbol{\alpha}'_J - \frac{\varepsilon'}{\delta'} \boldsymbol{\beta}'_J \\ \boldsymbol{\beta}_J = \frac{1}{\delta'} \boldsymbol{\beta}'_J \end{cases} \quad (3.2.4)$$

concluindo-se que τ goza da propriedade simétrica.

Por fim, admitamos que $(\mathbf{x}_b^i, \boldsymbol{\alpha}_J^i, \boldsymbol{\beta}_J^i) \tau (\mathbf{x}_b^{i+1}, \boldsymbol{\alpha}_J^{i+1}, \boldsymbol{\beta}_J^{i+1})$, existindo $\delta_i \neq 0$ e ε_i tais que (3.2.3) se verifica quando $i = 1, 2$. Sendo agora $\delta = \delta_1 \delta_2 \neq 0$ e $\varepsilon = \varepsilon_1 \delta_2 + \varepsilon_2$, as expressões (3.2.3) verificam-se para o par $[(\mathbf{x}_b^1, \boldsymbol{\alpha}_J^1, \boldsymbol{\beta}_J^1), (\mathbf{x}_b^3, \boldsymbol{\alpha}_J^3, \boldsymbol{\beta}_J^3)]$ o que mostra que τ goza da propriedade transitiva.

Por conseguinte fica demonstrado o lema. ■

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

Neste seguimento Pinto (2006) estabeleceu que, para a função objectivo, foram criados o gradiente e a matriz Hessiana de $S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J)$ (ver Pinto (2006), págs 14 a 17). Pereira (2004) efectuou uma simplificação da matriz Hessiana (ver , sendo que a mesma simplificação tem como base a suposição de que após algumas iterações do

algoritmo Zig-Zag se obtêm um bom ajustamento, sendo deste modo obtidos os estimadores dos índices ambientais $\tilde{x}_i, i = 1, \dots, b$ e os coeficientes de regressão ajustados $(\tilde{\alpha}_j, \tilde{\beta}_j), j = 1, \dots, J$.

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

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

serão bastante próximos do zero.

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

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

Seguidamente demonstra-se que a matriz D é definida positiva, o que pressupõe um bom desempenho do método ACR.

Sendo $\mathbf{u}'_{b+2,J} = [\mathbf{g}'_b, \mathbf{t}'_J, \mathbf{s}'_J]$, virá a forma quadrática

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

Concluindo-se que numa vizinhança do óptimo podemos considerar a matriz $Hess(S)$ (ver Oliveira (2007), pág. 18) definida positiva e nesse caso devemos esperar um bom comportamento do algoritmo Zig-Zag.

3.2.2. Redes de ensaios conectadas

Na mesma análise ACR, podemos considerar uma rede de ensaios como um único ensaio sendo que, o mesmo pode ser, ou não, conectado.

Se o ensaio não for conectado pode ser decomposto em sub-ensaios conectados e por conseguinte podemos dizer que a rede de ensaios foi decomposta em sub-redes conectadas.

Neste seguimento, na minimização descrita no algoritmo Zig-Zag, obtêm-se uma sequência de mínimos, existindo um mínimo absoluto se a rede for conectada. De acordo com Pereira (2004) e Pinto (2006) podemos estabelecer a seguinte proposição

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

Dem. *Se uma rede se divide em duas sub-redes conectadas, caso as mesmas tenham um ou mais cultivares em comum a rede global será conectada, ver Bose (1950), Raghavarao (1971) e Kageyama (1984). Assim é necessário e suficiente que os conjuntos de cultivares presentes nas duas sub-redes conectadas sejam disjuntos, para que a rede global não seja conectada. Sendo que este resultado se estende directamente às situações em que a rede global se decompõe em $L \geq 2$ sub-redes conectadas*

Sejam $B_l, l = 1, \dots, L$ os conjuntos dos índices dos blocos que integram cada uma das L sub-redes e $C_l, l = 1, \dots, L$ o conjunto dos índices de cultivares presentes na l -ésima sub-rede. Tem-se a função objectivo

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

O valor mínimo tomado por S será a soma de valores mínimos para as subfunções objectivo correspondentes às sub-redes. Assim, cada mínimo $(\tilde{\mathbf{x}}_b, \tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\beta}}_j)$ de S será a reunião de mínimos $(\tilde{\mathbf{x}}_{l,b_l}, \tilde{\boldsymbol{\alpha}}_{l,J_l}, \tilde{\boldsymbol{\beta}}_{l,J_l})$ das $S_l, l = 1, \dots, L$.

Por conseguinte podemos ver que se podem ter mínimos de S que não são equivalentes τ , estabelecendo desta forma a tese. ■

Devemos ter em atenção que a conexão na rede de ensaios é apenas a condição necessária para os mínimos de S sejam equivalentes τ . Para o mesmo facto apresentamos a seguinte proposição

Proposição 3.4. *Para que os mínimos sejam equivalentes τ é necessário e suficiente que,*

1. *A rede seja conectada*
2. $J \leq \text{Int} \left[\frac{n-b+2}{2} \right]$

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

Dem. *Atendendo à proposição 3.3 verifica-se que a condição 1 é necessária para que os mínimos de S sejam equivalentes τ . No entanto, mesmo que uma rede seja conectada existe ainda a possibilidade de o mínimo de S conter mais do que uma classe de equivalência τ . Por conseguinte, uma rede com n parcelas agrupadas em b blocos, com J cultivares para comparar, terá b índices ambientais e $2J$ coeficientes a estimar. Da proposição 3.1 sabe-se que se podem escolher os mínimos para que os mesmos estejam em \mathbf{V} , ou seja, introduzem-se duas condições relativas ao vector dos índices ambientais. Temos assim $b + 2J - 2$ parametros a ajustar, logo para que se possa ajustar sem ambiguidade o modelo devemos ter $n \geq b + 2J - 2$, o que verifica a condição 2, completando assim a demonstração. ■*

De acordo com Pinto (2006), dada uma rede conectada que satisfaça a segunda condição da proposição 3.4, estamos em condições de proceder ao ajustamento utilizando o algoritmo Zig-Zag.

De acordo com Oliveira (2007), para se efectuar uma análise global duma rede de ensaios é exigível que a mesma seja conectada. Caso a mesma não o seja, devemos decompor em sub-redes conectadas maximais e proceder a uma análise individual.

3.2.3. Algoritmo Zig-Zag

Como referimos acima, o algoritmo Zig-Zag 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.

Este mesmo algoritmo é frequentemente usado no ajustamento de regressões lineares dos rendimentos nos índices ambientais, apresentando na grande maioria dos casos um bom desempenho.

Neste ponto pretendemos realizar uma descrição detalhada do algoritmo e apresentar algumas das justificações que nos permitem aferir sobre uma boa qualidade dos ajustamentos.

Tal como referimos desde o início do presente capítulo, pretendemos minimizar a função objectivo

$$S(\mathbf{x}_b, \boldsymbol{\alpha}_J, \boldsymbol{\beta}_J) = \sum_{j=1}^J \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2 \quad (3.2.1)$$

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

Quando dispomos de ensaios com blocos completos, consideram-se os rendimentos médios dos blocos, como valores iniciais para os índices ambientais.

Se tivermos um delineamento do tipo α -design, ver Patterson & Williams (1976), os blocos estão agrupados em superblocos, onde cada cultivar estará representado α vezes. Nesta situação atribui-se a cada bloco, o rendimento médio para o superbloco a que pertence, como valor inicial para o respectivo índice ambiental.

Tomando como opção considerar os rendimentos médios dos blocos com valores iniciais para os respectivos índices ambientais, e sendo $x_i^0, i = 1, \dots, b$, os respectivos valores iniciais, começaremos por minimizar

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

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

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

Seguidamente procedemos à minimização do vector x_b , sendo que neste estudo pretendemos minimizar a função objectivo

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

com

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

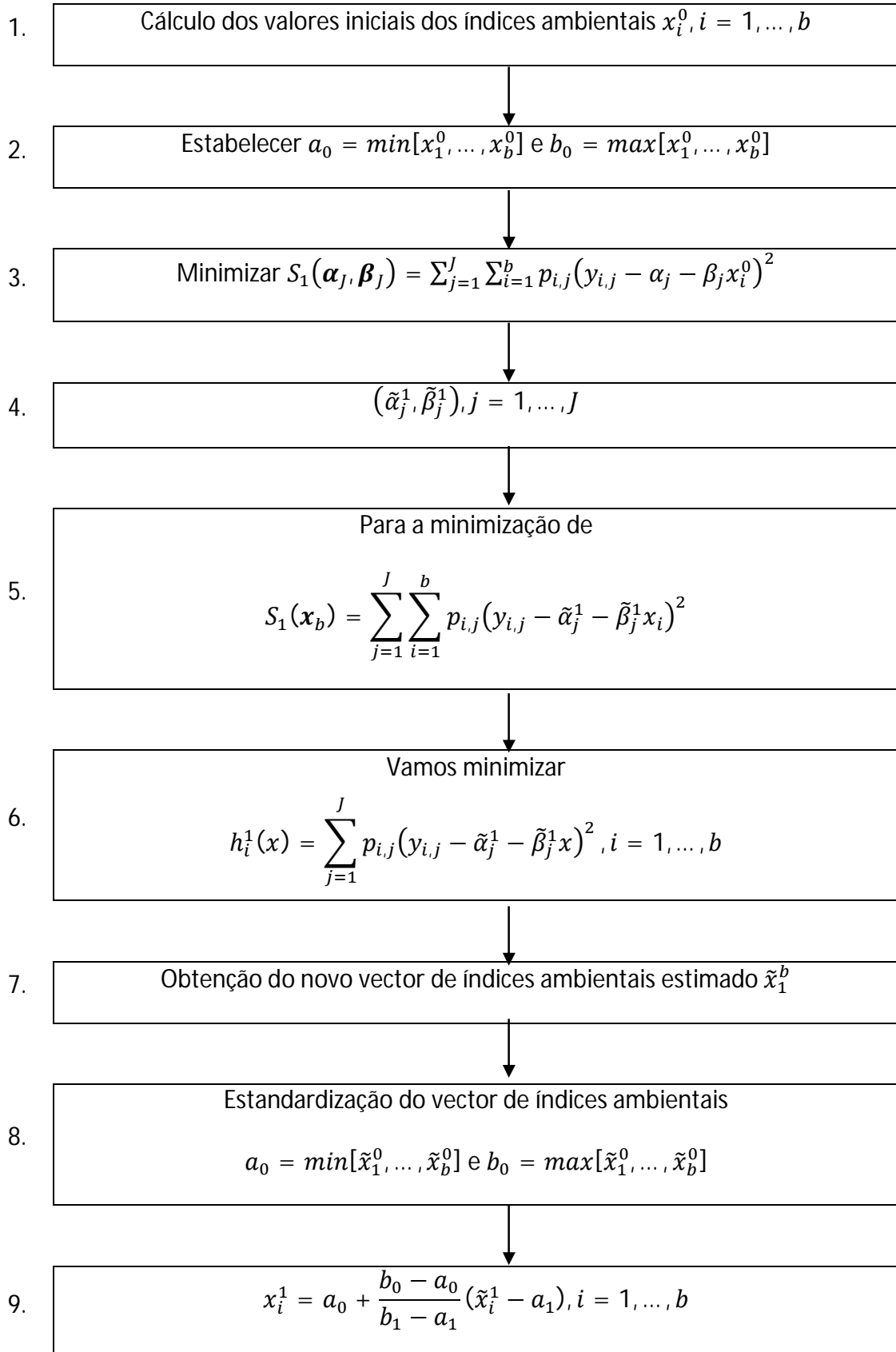
sendo as componentes dos mínimos dadas por

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

repetindo-se as iterações até a soma dos quadrados dos resíduos estabilizar.

Seguidamente podemos visualizar uma esquematização do algoritmo Zig-Zag.

Esquema 1. Algoritmo Zig-Zag



Devemos repetir o procedimento a partir do segundo passo (Estabelecer $a_0 = \min[x_1^0, \dots, x_b^0]$ e $b_0 = \max[x_1^0, \dots, x_b^0]$) 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 conectada como foi descrito anteriormente na proposição 3.3.

Pereira (2004) refere que para os mínimos de S sejam equivalentes τ é necessário e suficiente que se verifique a proposição 3.4 e deste modo estamos em condições para construir o algoritmo para obtermos o mínimo $(\tilde{x}_b, \tilde{\alpha}_j, \tilde{\beta}_j)$ da função objectivo $S(x_b, \alpha_j, \beta_j)$.

3.3. Caso completo

De acordo com Oliveira (2007), no início a ACR foi aplicada a redes de ensaios em blocos casualizados. Neste contexto em Gusmão (1985), (1986a) e (1986b) são utilizados os índices ambientais correspondentes aos diferentes blocos medidos pelos respectivos rendimentos médios, mostrando-se que a precisão das análises dos delineamentos em blocos casualizados aumenta de modo significativo se foram considerados índices ambientais por blocos em vez de um só índice ambiental por ensaio, explicando-se que tal se deve ao facto de existir um aumento do número de pontos obtidos por regressão.

Neste caso a técnica clássica ajusta cultivar a cultivar regressões lineares das produções nas médias obtidas para cada bloco. Este tipo de procedimento, segundo Oliveira (2007) tem duas contudo duas desvantagens:

- i. Não ser de todo correcto 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 Zig-Zag nesta técnica, algoritmo descrito acima, ou a técnica de dupla minimização, técnica mais robusta.

3.3.1. Técnica de Dupla Minimização

Considerando que todos os cultivares estão presentes em todos os blocos, então a função objectivo é a seguinte

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

onde os pesos p_i correspondem a blocos do mesmo ensaio.

Consideremos as somas

$$T_x = \sum_{i=1}^b p_i x_i \quad ; \quad T_{iy} = \sum_{i=1}^b p_i y_{i,j} \quad ; \quad j = 1, \dots, J$$

$$S_{xx} = \sum_{i=1}^b p_i x_i^2 \quad ; \quad S_{jxy} = \sum_{i=1}^b p_i x_i y_{i,j} \quad ; \quad S_{jyy} = \sum_{i=1}^b p_i y_{i,j}^2$$

e as somas corrigidas

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

$$s_{jxy} = S_{jxy} - \frac{1}{\sum_{i=1}^b p_i} \sum_{i=1}^b p_i x_i \sum_{i=1}^b p_i y_{i,j}$$

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

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

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

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

$$\tilde{\alpha}_j(\mathbf{x}_b) = \frac{1}{\sum_{i=1}^b p_i} (T_{jy} - \tilde{\beta}_j(\mathbf{x}_b) T_x)$$

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

Por conseguinte o mínimo condicional para a soma dos quadrados dos resíduos será

$$\tilde{S}(\mathbf{x}_b) = S(\mathbf{x}_b, \tilde{\alpha}_j(\mathbf{x}_b), \tilde{\beta}_j(\mathbf{x}_b))$$

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

Para minimizarmos $\tilde{S}(\mathbf{x}_b)$ dado que $\sum_{j=1}^J s_{jyy}$ não depende de \mathbf{x}_b , temos que maximizar $\sum_{j=1}^J \frac{s_{jxy}^2}{s_{xx}}$

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

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

$$= \sum_{j=1}^J \left(\sum_{i=1}^b p_i x_i \left(y_{i,j} - \frac{\sum_{i=1}^b p_i y_{i,j}}{\sum_{i=1}^b p_i} \right) \right)^2$$

Verificando-se

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

E sendo \mathbf{Y}_j^{+b} o vector de componentes

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

temos

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

Proposição 3.5. O vector x_b que satisfaz as restrições $\sum_{i=1}^b p_i x_i = 0$ e $\sum_{i=1}^b p_i x_i^2 = 1$, maximiza $M(x_b)$ é o primeiro vector próprio da matriz $(\sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j^{+b})$.

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

$$\left(\sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j^{+b} \right) \left(\frac{1}{b^2} \mathbf{1}_b \right) = \sum_{j=1}^J \mathbf{Y}_j^{+b} \left(\mathbf{Y}_j^{+b} \frac{1}{b^2} \mathbf{1}_b \right) = 0$$

Por conseguinte $\left(\frac{1}{b^2} \mathbf{1}_b \right)$ será um vector próprio de $(\sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j^{+b})$ estando-lhe associado um valor próprio nulo.

Por outro lado, tendo-se

$$\mathbf{u}'_b \sum_{j=1}^J \mathbf{Y}_j^{+b} \mathbf{Y}_j^{+b} \mathbf{u}_b = \sum_{j=1}^J (\mathbf{u}'_b \mathbf{Y}_j^{+b}) (\mathbf{Y}_j^{+b} \mathbf{u}_b) = \sum_{j=1}^J (\mathbf{u}'_b \mathbf{Y}_j^{+b})^2 \geq 0$$

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

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

Este mesmo vector próprio $\mu_{1,b}$ terá de ser ortogonal a $\left(\frac{1}{b^2} \mathbf{1}_b \right)$, tendo soma das componentes nula (ver Giraldez et al., 1995).

Por conseguinte a demonstração fica concluída se tivermos conta que $\|\mu_{1,b}\| = 1$ e dado que $\mu_{1,b}$ satisfaz as restrições $\sum_{i=1}^b p_i x_i = 0$ e $\sum_{i=1}^b p_i x_i^2 = 1$ e maximiza $M(x_b)$ na família de todos os vectores. ■

3.3.2. Estimadores de Máxima Verosimilhança

De entre vários procedimentos tendo com finalidade a obtenção dos melhores estimadores, podemos destacar vários métodos, o método dos mínimos quadrados, o método dos momentos e o método da máxima verosimilhança são três dos métodos mais conhecidos e mais aplicados, o último tem uma grande aplicabilidade na inferência estatística.

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

Deste modo a função de verosimilhança será a seguinte

$$L(\alpha_J, \beta_J, x_b, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j \in D_J} p_i (y_{i,j} - \alpha_j - \beta_j x_i)^2}}{(2\pi)^{\frac{N}{2}} \sigma^N}$$

Logaritimizando a mesma obtemos

$$\ln(L(\alpha_J, \beta_J, x_b, \sigma^2)) = -\frac{1}{2\sigma^2} \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2 + C$$

onde

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

Começando por deduzir o estimador de σ^2 fazendo

$$\frac{\partial \ln(L(\alpha_J, \beta_J, x_b, \sigma^2))}{\partial \sigma^2} = \frac{1}{2\sigma^4} \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2 - \frac{N}{2\sigma^2}$$

e igualando a zero obtemos

$$\tilde{\sigma}^2 = \frac{1}{N} \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (y_{i,j} - \alpha_j - \beta_j x_i)^2$$

Para obtermos $(\tilde{\alpha}_j, \tilde{\beta}_j), j = 1, \dots, J$ e $\tilde{x}_i, i = 1, \dots, b$ teremos que minimizar

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

Sendo assumida a normalidade, os estimadores de mínimos quadrados são também estimadores de máxima verosimilhança (ver Oliveira, 2007).

3.3.3. Testes de razão de verosimilhanças

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

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 de maximizar

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

Calculando

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

E igualando a zero, teremos

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

Para se obter $(\tilde{\alpha}_j, \tilde{\beta}_j), j = 1, \dots, J$ e $\tilde{x}_i, i = 1, \dots, b$ vamos minimizar

$$S_w(\alpha_j, \beta, x_b) = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (y_{i,j} - \alpha_{jw} - \beta_w \tilde{x}_{i,w})^2$$

Recorrendo ao algoritmo Zig-Zag, minimiza-se em cada iteração, primeiro em onde a (α_j, β_j) e seguidamente em onde a x_b .

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

Derivando a função S_w em ordem aos coeficientes obtemos

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

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

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

pelo que

$$\beta_w = \frac{\sum_{i=1}^b p_{i,j} y_{i,j} - \sum_{i=1}^b p_{i,j} \alpha_{j,w}}{\sum_{i=1}^b p_{i,j} x_{i,w}}$$

com $\alpha_{j,w} = \alpha_1 + d_{j,w}, j = 2, \dots, J$

sendo

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

Igualando também a zero as derivadas $\frac{\partial S_w}{\partial \beta}$, vem

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

Consequentemente

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

De ambas as expressões para β_w , podemos concluir que $\alpha_1 = \frac{C_1 - C_2}{D_1 - D_2}$, onde

$$\begin{cases} C_1 = \frac{\sum_{i=1}^b p_{i,j} \sum_{j=1}^J y_{i,j} - \sum_{i=1}^b p_{i,j} \sum_{j=1}^J d_{j,w}}{J \sum_{i=1}^b p_{i,j} x_{i,w}} \\ C_2 = \frac{\sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,w} y_{i,j} - \sum_{i=1}^b p_{i,j} x_{i,w} \sum_{j=1}^J d_{j,w}}{J \sum_{i=1}^b p_{i,j} x_{i,w}^2} \\ D_1 = \frac{\sum_{i=1}^b p_{i,j}}{\sum_{i=1}^b p_{i,j} x_{i,w}} \\ D_2 = \frac{\sum_{i=1}^b p_{i,j} x_{i,w}}{\sum_{i=1}^b p_{i,j} x_{i,w}^2} \end{cases}$$

Daqui obtêm-se os $\alpha_2, \dots, \alpha_j$ e o β correspondentes à primeira iteração passando-se seguidamente à maximização em ordem aos índices ambientais.

A função a maximizar é a seguinte

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

Derivando em ordem a x_i e igualando a zero vem

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

e logo

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

Seguidamente passando a teste a hipótese

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

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

Logo

$$\tilde{\sigma}_\omega^2 = \frac{1}{bJ} \sum_{i=1}^b p_{i,j} \sum_{j=1}^J (y_{i,j} - \tilde{\alpha}_\omega - \tilde{\beta}_{j,\omega} \tilde{x}_{i,\omega})^2$$

Por conseguinte devemos minimizar recorrendo ao algoritmo de Zig-Zag a função

$$S_\omega(\alpha, \beta_j, x_b) = \sum_{i=1}^b \sum_{j=1}^J p_{i,j} (y_{i,j} - \alpha_\omega - \beta_{j,\omega} x_{i,\omega})^2$$

para obter $\tilde{\alpha}_\omega$, $\tilde{\beta}_{j,\omega}$ e $\tilde{x}_{b,\omega}$.

Seja

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

com

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

vindo

$$\left\{ \begin{array}{l} C_1 = \frac{\sum_{i=1}^b p_{i,j} x_{i,\omega} \sum_{j=1}^J y_{i,j} - \sum_{i=1}^b p_{i,j} x_{i,\omega}^2 \sum_{j=1}^J d_{j,\omega}}{J \sum_{i=1}^b p_{i,j} x_{i,\omega}} \\ C_2 = \frac{\sum_{i=1}^b p_{i,j} \sum_{j=1}^J y_{i,j} - \sum_{i=1}^b p_{i,j} x_{i,\omega} \sum_{j=1}^J d_{j,\omega}}{J \sum_{i=1}^b p_{i,j}} \\ D_1 = \frac{\sum_{i=1}^b p_{i,j} x_{i,\omega}^2}{\sum_{i=1}^b p_{i,j} x_{i,\omega}} \\ D_2 = \frac{\sum_{i=1}^b p_{i,j} x_{i,\omega}}{\sum_{i=1}^b p_{i,j}} \end{array} \right.$$

Por conseguinte, na maximização em ordem aos índices ambientais pode tomar-se

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

Por fim, devemos testar

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

Neste teste

$$\tilde{\sigma}_\omega^2 = \frac{1}{bJ} \sum_{i=1}^b p_{i,j} \sum_{j=1}^J (y_{i,j} - \tilde{\alpha}_\omega - \tilde{\beta}_{j,\omega} \tilde{x}_{i,\omega})^2$$

E vamos minimizar

$$S_\omega(\alpha, \beta, x_b) = \sum_{i=1}^b p_{i,j} \sum_{j=1}^J (y_{i,j} - \alpha - \beta x_{i,\omega})^2$$

Derivando S_ω em ordem aos coeficientes α e β obtemos

$$\begin{cases} \frac{\partial S_\omega}{\partial \alpha} = -2 \sum_{i=1}^b p_{i,j} (y_{i,j} - \alpha - \beta x_{i,\omega}) \\ \frac{\partial S_\omega}{\partial \beta} = -2 \sum_{i=1}^b \sum_{j=1}^J p_{i,j} x_{i,\omega} (y_{i,j} - \alpha - \beta x_{i,\omega}) \end{cases} \quad , j = 1, \dots, J$$

Igualando as derivadas a zero obtemos um sistema do tipo

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

cuja solução será

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

Neste caso, tem-se

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

$$b_1 = \sum_{i=1}^b p_{i,j} \sum_{j=1}^J y_{i,j} \quad ; \quad b_2 = \sum_{i=1}^b p_{i,j} x_{i,\omega} \sum_{j=1}^J y_{i,j}$$

vindo

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

Sendo que na segunda parte da iteração obtêm-se directamente

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

3.4. A importância da Estabilidade dos Cultivares

De acordo com Oliveira (2007), a existência de estabilidade no comportamento dos cultivares é traduzida pela ausência de diferenças significativas nos coeficientes angulares das regressões ajustadas.

De acordo com o mesmo autor, a mesma característica tem grande importância na criação de um plano de melhoramento onde é valorizada a obtenção de cultivares que tenham boa estabilidade de resposta às variações dos índices ambientais.

Por conseguinte para verificarmos a existência de estabilidade dos cultivares ao longo dos anos apresenta-se o seguinte teste

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

onde $\{i_1, \dots, i_j\}$ representa o conjunto dos índices relativos aos anos em que o j -ésimo cultivar foi utilizado.

Para construirmos a estatística de teste vamos estabelecer o seguinte lema

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

Dem. *Seja \mathbf{K} uma matriz definida positiva, podemos escrever $\mathbf{K} = \mathbf{P}'\mathbf{D}(r_1, \dots, r_m)\mathbf{P}$ com \mathbf{P} matriz ortogonal $r_i > 0, i = 1, \dots, m$ os valores próprios positivos de \mathbf{K} . A matriz*

$\mathbf{K}^{\frac{1}{2}} = \mathbf{P}'\mathbf{D}\left(r_1^{\frac{1}{2}}, \dots, r_m^{\frac{1}{2}}\right)\mathbf{P}$ também é definida positiva e tem-se $\mathbf{K}^{\frac{1}{2}}\mathbf{K}^{\frac{1}{2}} = \mathbf{K}$. Deste modo temos $\mathbf{AKA}' = \mathbf{AK}^{\frac{1}{2}}\mathbf{A}' = \left(\mathbf{AK}^{\frac{1}{2}}\right)\left(\mathbf{AK}^{\frac{1}{2}}\right)'$ e como $\mathbf{K}^{\frac{1}{2}}$ é regular ter-se-á, ser Silvey (1975), $\text{car}\left(\mathbf{AK}^{\frac{1}{2}}\right) = \text{car}(\mathbf{AKA}') = \text{car}(\mathbf{A}) = m - 1$ pelo que \mathbf{AKA}' será regular e sendo para $\forall \omega \neq 0, \omega'(\mathbf{AKA}')\omega = \left\|\mathbf{AK}^{\frac{1}{2}}\omega\right\|^2 > 0$, então \mathbf{AKA}' será definida positiva. Por conseguinte a demonstração fica completa observando-se que se uma matriz é definida positiva a sua inversa também o é. ■

Estabeleçamos seguidamente

Proposição 3.8. *Seja $\underline{\mathbf{Z}}_m \sim N\left(\underline{\boldsymbol{\eta}}_m, \sigma^2 \mathbf{K}\right)$ com \mathbf{K} matriz definida positiva, independente de $S \sim \sigma^2 \chi_g^2$ e \mathbf{A} a matriz obtida no lema anterior.*

A estatística de teste é

$$F = \frac{g \left(\underline{\mathbf{AZ}}_m\right)' \left(\mathbf{AKA}'\right)^{-1} \left(\underline{\mathbf{AZ}}_m\right)}{r S}$$

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

Dem. Em Seber (1980, pág. 5), podemos ver que $\underline{\mathbf{AZ}}_m \sim N\left(\mathbf{A}\underline{\boldsymbol{\eta}}_m, \sigma^2 \mathbf{AKA}'\right)$, pelo que, ver Mexia (1995, pág. 52), $\left(\underline{\mathbf{AZ}}_m\right)' \left(\mathbf{AKA}'\right)^{-1} \left(\underline{\mathbf{AZ}}_m\right) \sim \sigma^2 \chi_{m-1, \delta}^2$ independente de $S \sim \sigma^2 \chi_g^2$ e por conseguinte a primeira parte está demonstrada. Recorrendo ao lema 2.1, a matriz \mathbf{AKA}' é definida positiva, pelo que podemos escrever $\mathbf{AKA}' = \mathbf{P}'\mathbf{D}(r_1, \dots, r_m)\mathbf{P}$, com \mathbf{P} matriz ortogonal. Sendo $r_i > 0, i = 1, \dots, r$, os valores próprios de \mathbf{AKA}' , ter-se-á então, $\left(\mathbf{AKA}'\right)^{-1} = \mathbf{P}'\mathbf{D}(r_1^{-1}, \dots, r_r^{-1})\mathbf{P}$, vindo $\delta = 0$ sse $\mathbf{PA}\underline{\boldsymbol{\eta}}_m = 0_r$, o que, por \mathbf{P} ser regular, equivale a ter-se $\mathbf{A}\underline{\boldsymbol{\eta}}_m = 0_r$. Ora, \mathbf{A} é constituída por r linhas de uma matriz ortogonal de ordem m , tendo-se $\text{car}(\mathbf{A}) = r$ e sendo $N(\mathbf{A})$ o espaço de nulidade de \mathbf{A} , $\text{dim}(N(\mathbf{A})) = m - \text{car}(\mathbf{A}) = 1$. Basta agora observar que por construção $\frac{1}{\sqrt{m}}\mathbf{1}_m$ é ortogonal aos vectores linha de \mathbf{A} , para se ver que os vectores

de $N(\mathbf{A})$ são do tipo $\frac{c}{\sqrt{m}} \mathbf{1}_m$. Teremos então, $\delta = 0$ sse $\underline{\boldsymbol{\eta}}_m = \frac{c}{\sqrt{m}} \mathbf{1}_m$ o que estabelece a segunda parte da demonstração. ■

Numa abordagem prática, quando dispomos de coeficientes ajustados em cada ano, ou noutro período temporal. Consequentemente sendo as somas dos quadrados dos desvios para a média dos índices ambientais ajustados nesse ano $s_{a_l}, a_l, l = 1, \dots, m_j$ é o índice correspondente, para o j -ésimo cultivar tomamos

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

e

$$\underline{\mathbf{z}}'_m = [\tilde{\beta}_{j,a_1}, \dots, \tilde{\beta}_{j,a_{m_j}}]$$

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

Capítulo 4.

Método do Contorno Superior

4.1. Selecção pelo método do Contorno Superior

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

De acordo com Pinto (2006), 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.

Começemos por considerar j regressões ajustadas

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

uma por cultivar e considerando $[\theta_{min}; \theta_{max}]$ o intervalo cujos extremos são exactamente o mínimo e o máximo dos índices ambientais ajustados.

De entre todas as regressões, teremos no intervalo $[\theta_{min}; \theta_{max}]$ as regressões dominantes que são aquelas que correspondem às produções máximas, logo integrando o contorno superior. Neste seguimento podemos considerar que os cultivares associados às regressões dominantes serão naturalmente os cultivares dominantes, como referimos acima.

De acordo com Oliveira (2007), na selecção dos cultivares a recomendar, deve obviamente ser dada principal atenção aos que correspondem às regressões que integram o contorno superior. Estes serão comparados com todos os restantes, e no caso de algum destes ser significativamente dominado, deverá de imediato ser eliminado do processo de selecção. Neste trabalho estamos particularmente interessados em desenvolver um processo automático para identificação dos cultivares que integram o contorno superior.

4.1.1. Geometria

Como estamos perante regressões ajustadas do tipo linear, o contorno superior será uma poligonal aberta composta por segmentos de recta.

Neste sentido, atribuindo índices j , ($j = 1, \dots, J$), a essas regressões ajustadas, e ordenando-as por ordem decrescente dos seus declives ficará

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

tendo-se o caso particular $\tilde{\alpha}_j > \tilde{\alpha}_{j+1}$ quando $\tilde{\beta}_j = \tilde{\beta}_{j+1}$

Segundo Oliveira (2007), se existirem apenas rectas paralelas a que tiver maior ordenada na origem fará parte do contorno superior e conseqüentemente apenas esse cultivar poderá ser dominante, por conseguinte os cultivares significativamente dominados deverão ser excluídos do ensaio.

Seguidamente iremos estabelecer alguns resultados que permitem caracterizar geometricamente o traçado do contorno superior. São apresentados exemplos em R para cada um dos casos possíveis.

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 rectas se intersectam, à direita do ponto de intersecção, fica por cima a recta 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. ■

Por outro lado, quando se têm $\tilde{\beta}_j \neq \tilde{\beta}_{j'}$,

$$\theta_{jj'} = \frac{\tilde{\alpha}_j - \tilde{\alpha}_{j'}}{\tilde{\beta}_{j'} - \tilde{\beta}_j}, j \neq j' e j, j' = 1, \dots, J$$

é o ponto de intersecção entre as rectas de índices j e j' e as coordenadas do mesmo ponto serão $(\tilde{\theta}_{jj'}, \tilde{\alpha}_j + \tilde{\beta}_j \tilde{\theta}_{jj'})$. Como podemos verificar na figura abaixo, quando

dispomos de dois cultivares onde as rectas de regressão linear simples têm declives diferentes, existe um ponto de intersecção entre as mesmas (Figura 4.1).

A mesma figura foi obtida com recurso ao programa desenvolvido em R designado por "uppercontour" que está no Anexo A e descrito no Capítulo 5. Para a mesma figura devemos indicar a seguinte intrução

```
> uppercontour(dados,"índice ambiental mínimo","índice ambiental máximo")
```

```
> dados<-read.table('clipboard',header=T)
```

```
> dados
```

```
declive ordenada
```

```
1      1.3      30
```

```
2      0.8      70
```

```
> dados_ord
```

```
declive ordenada
```

```
1      1.3      30
```

```
2      0.8      70
```

```
> uppercontour(dados,0,300)
```

```
O cultivar 1 é dominante entre 80 e 300
```

```
O cultivar 2 é dominante entre 0 e 80
```

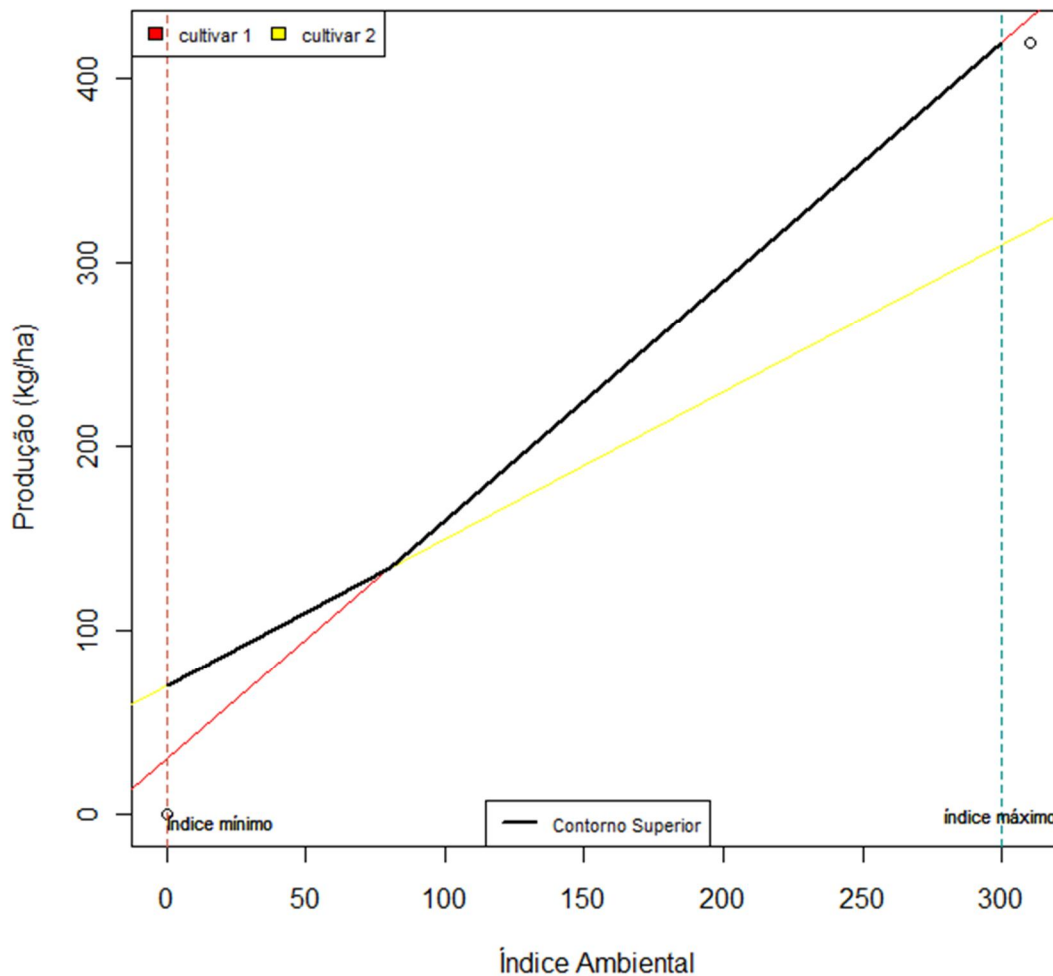


Figura 4.1. Traçado do contorno superior para duas rectas com declives diferentes.

Se não existir nenhum ponto de intersecção entre as duas rectas com índices j e j' as mesmas são consideradas paralelas, facto que podemos verificar através da figura abaixo onde as rectas são paralelas mas a recta a vermelho é a dominante pois tem maior ordenada na origem (Figura 4.2).

```
> dados<-read.table("C:\\dados.txt",header=T)
> dados
  ordenada declive
1       30      1.1
```

```
2      70      1.1
```

```
> dados_ord
```

```
ordenada declive
```

```
1      70      1.1
```

```
2      30      1.1
```

```
> uppercontour(dados, 0, 300)
```

O cultivar 1 é dominante entre 0 e 300

Os restantes cultivares não fazem parte do contorno superior!

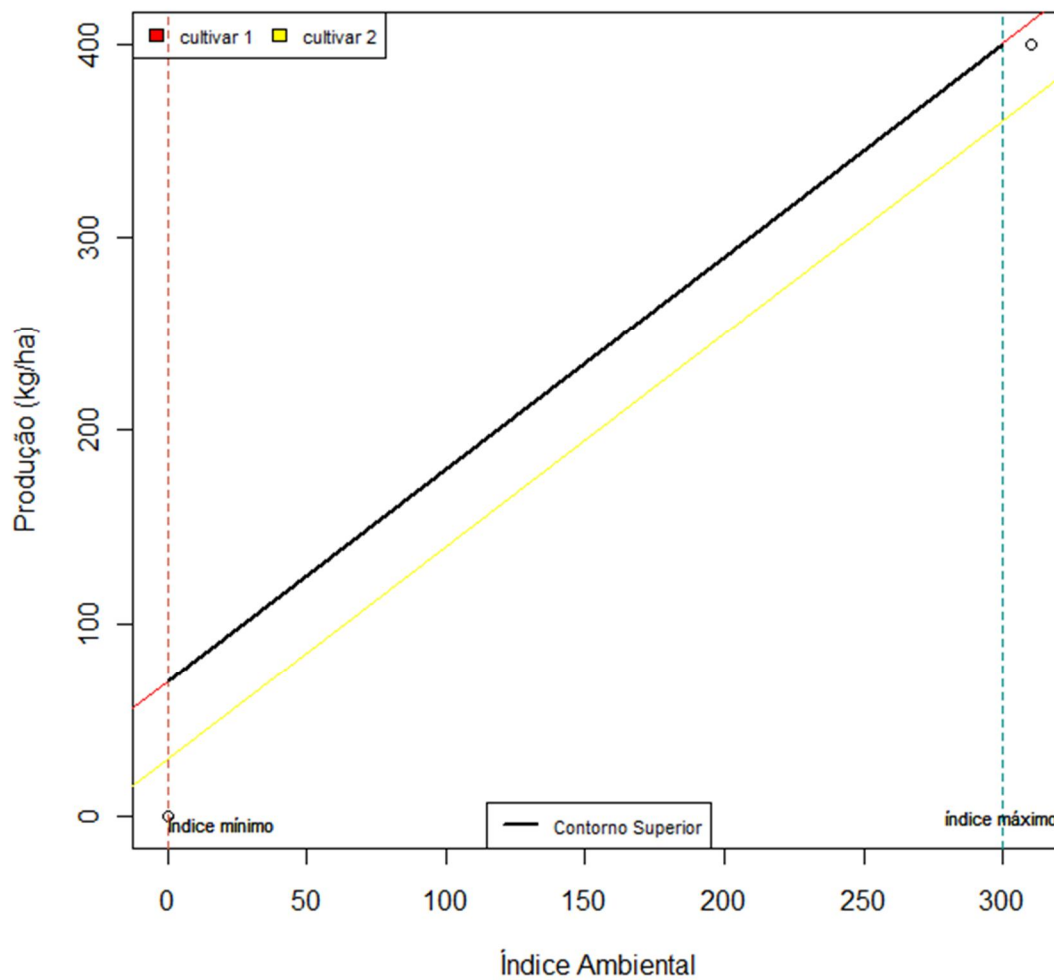
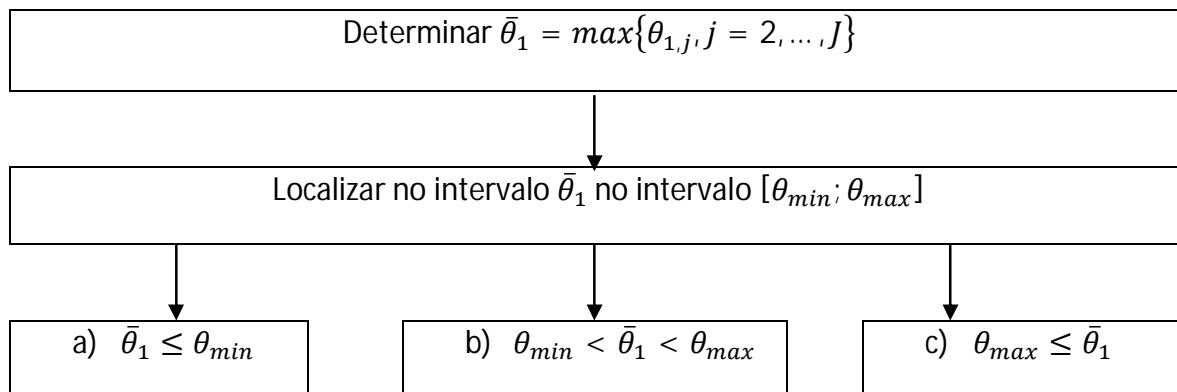


Figura 4.2. Traçado do contorno superior para duas rectas com declives iguais.

Para construir a poligonal que forma o contorno superior, podemos seguir o seguinte esquema

Esquema 2. Determinação do Contorno Superior



De acordo com o Esquema 2, podemos encontrar três situações

a) $\bar{\theta}_1 \leq \theta_{min}$

Neste caso, o contorno superior reduz-se apenas à recta ajustada $\tilde{\alpha}_1 + \tilde{\beta}_1 x$ com índice 1, sendo que apenas o primeiro cultivar será dominante . (Figura 4.3)

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

```
> dados
```

```
ordenada declive
```

```
1      40      0.8
2      30      1.3
3      35      1.2
4      42      1.5
```

```
> dados_ord
```

```
ordenada declive
```

```
4      42      1.5
2      30      1.3
3      35      1.2
```

1 40 0.8

> uppercontour(dados, 0, 300)

O cultivar 1 é dominante entre 0 e 300

Os restantes cultivares não fazem parte do contorno superior!

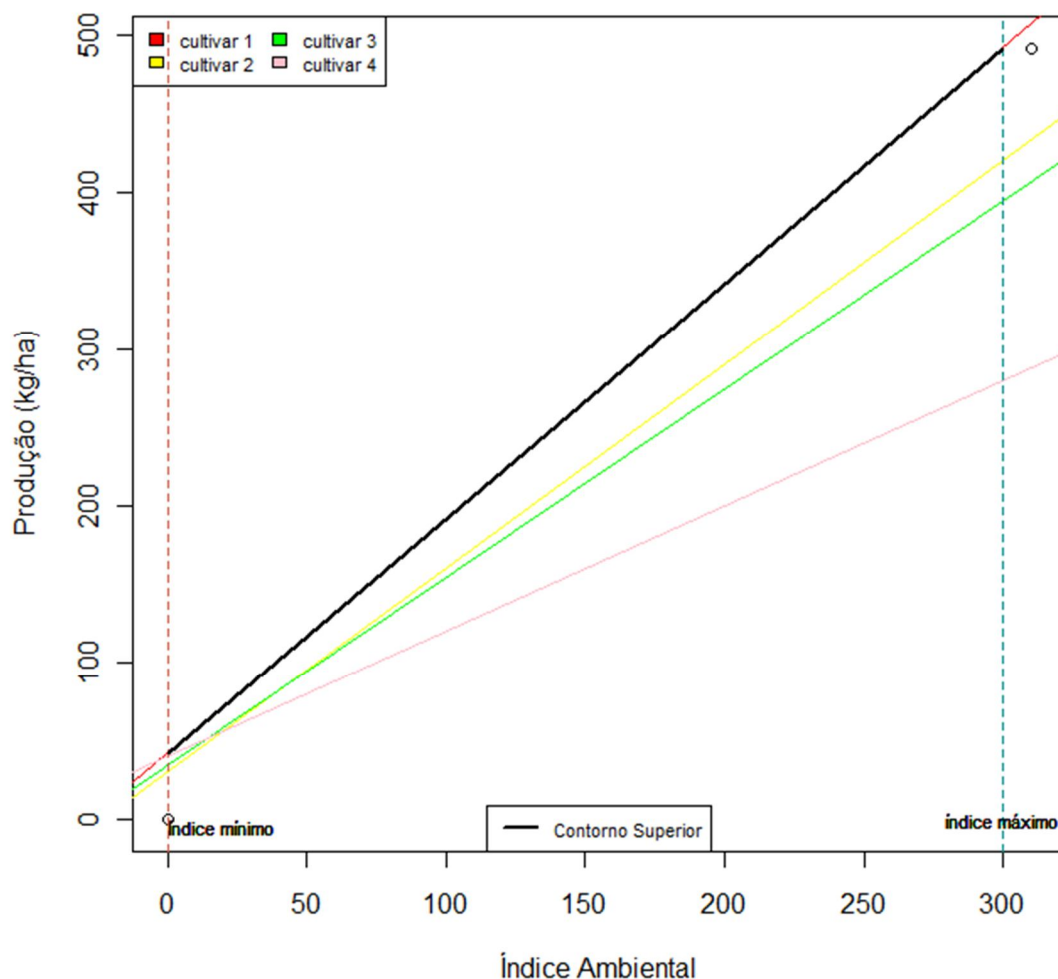


Figura 4.3. Traçado do contorno superior para quando $\bar{\theta}_1 \leq \theta_{min}$.

b) $\theta_{min} < \bar{\theta}_1 < \theta_{max}$

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

é dominante e para se averiguar da existência de mais cultivares dominantes terá de se obter

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

Neste seguimento para o intervalo $[\theta_{min}, \bar{\theta}_2]$ terá de se proceder do mesmo modo para determinar $\bar{\theta}_3$ e assim sucessivamente até que se tenha o contorno superior totalmente definido entre θ_{min} e θ_{max} (Figura 4.4)

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

```
> dados
```

```
ordenada declive
1          50      0.8
2          30      0.9
3          35      0.6
4          20      1.8
```

```
> dados_ord
```

```
ordenada declive
4          20      1.8
2          30      0.9
1          50      0.8
3          35      0.6
```

```
> uppercontour(dados,0,300)
```

```
O cultivar 1 é dominante entre 30 e 300
```

```
O cultivar 3 é dominante entre 0 e 30
```

```
Os restantes cultivares não fazem parte do contorno superior!
```

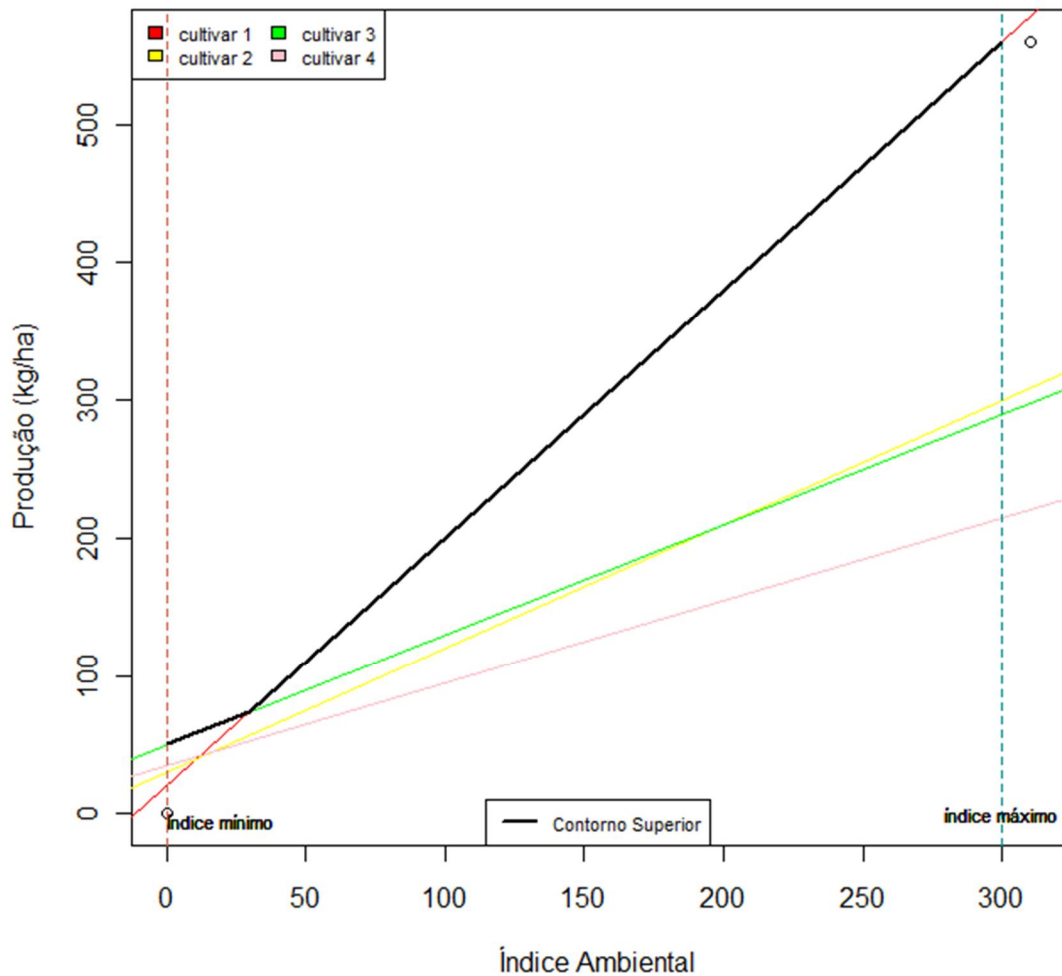


Figura 4.4. Traçado do contorno superior para quando $\theta_{min} < \bar{\theta}_1 < \theta_{max}$

c) $\theta_{max} \leq \bar{\theta}_1$

Neste mesmo caso a recta ajustada com índice 1 não faz parte do contorno superior e por conseguinte concluímos que o cultivar não é dominante para o ensaio realizado. (Figura 4.5)

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

> dados
  declive ordenada
1     1.2         50
2     1.3         30
3     1.6         35
4     1.8         10

> dados_ord
  declive ordenada
4     1.8         10
3     1.6         35
2     1.3         30
1     1.2         50

> uppercontour(dados,0,100)

[1] "O cultivar 1 não faz parte de contorno superior"
    O cultivar 2 é dominante entre 37.5 e 100
    O cultivar 4 é dominante entre 0 e 37.5
    Os restantes cultivares não fazem parte do contorno
superior!
```

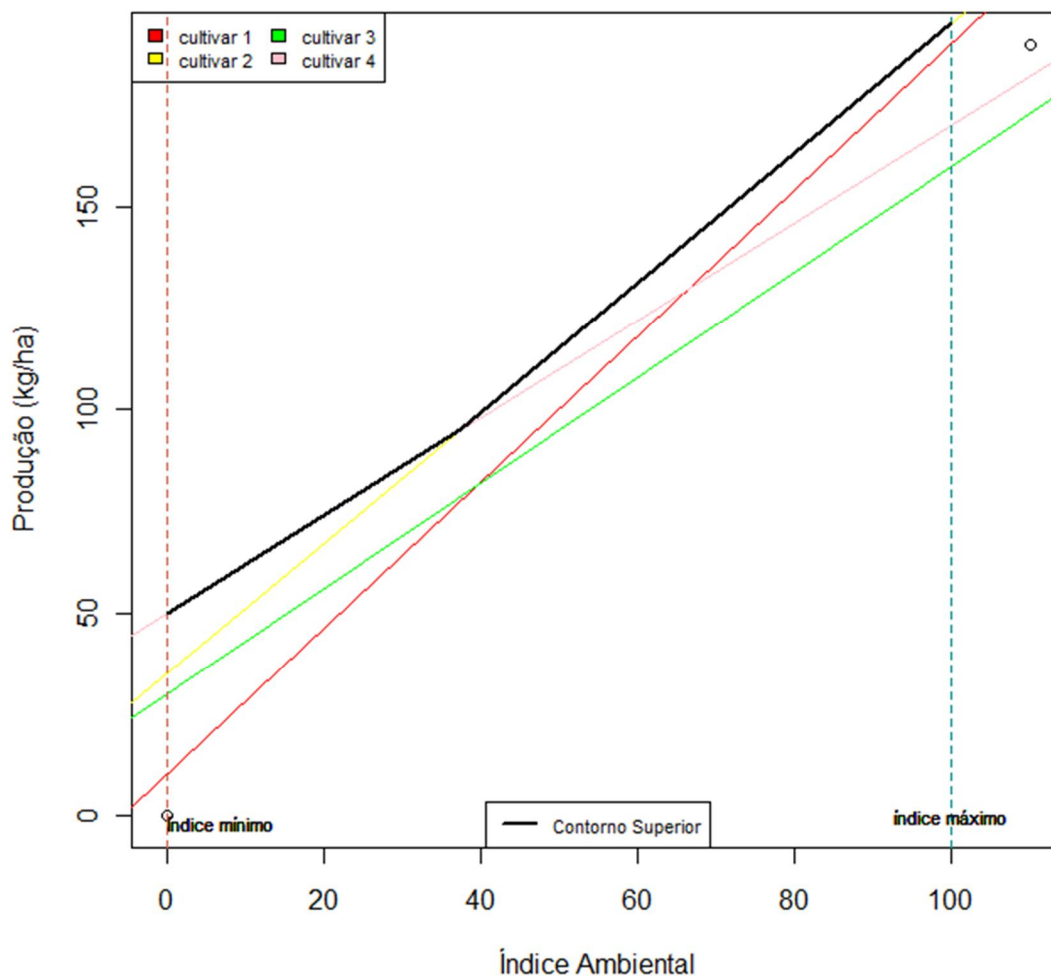


Figura 4.5. Traçado do contorno superior para quando $\theta_{max} \leq \bar{\theta}_1$

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

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

e verificar se:

- i. Caso $\theta_{max} \leq \bar{\theta}_2$, o cultivar com índice 2 não faz parte do contorno superior a passa-se a trabalhar com as rectas de índices $j = 3, \dots, J$
- ii. Caso $\bar{\theta}_2 \in [\theta_{min}; \theta_{max}]$ estamos numa situação idêntica à descrita na alínea b).

- iii. Devemos realizar o processo até que o contorno superior esteja definido, recorrendo às três alíneas a), b) e c). (Figura 4.6)

```
> dados<-read.table("C:\\dados.txt",header=T)
> dados
  declive ordenada
1     1.2     -40
2     1.5      38
3     1.8    -36
4     1.8     10
5     2.3    -50
6     1.1    -35
7     0.8     42
> dados_ord<-dados[order(-dados$declive,-dados$ordenada),]
> dados_ord
  declive ordenada
5     2.3    -50
4     1.8     10
3     1.8    -36
2     1.5     38
1     1.2    -40
6     1.1    -35
7     0.8     42
> uppercontour(dados,0,150)
O cultivar 1 é dominante entre 120 e 150
O cultivar 2 é dominante entre 93.33333 e 120
O cultivar 4 é dominante entre 5.714286 e 93.33333
O cultivar 7 é dominante entre 0 e 5.714286
```

Os restantes cultivares não fazem parte do contorno superior!

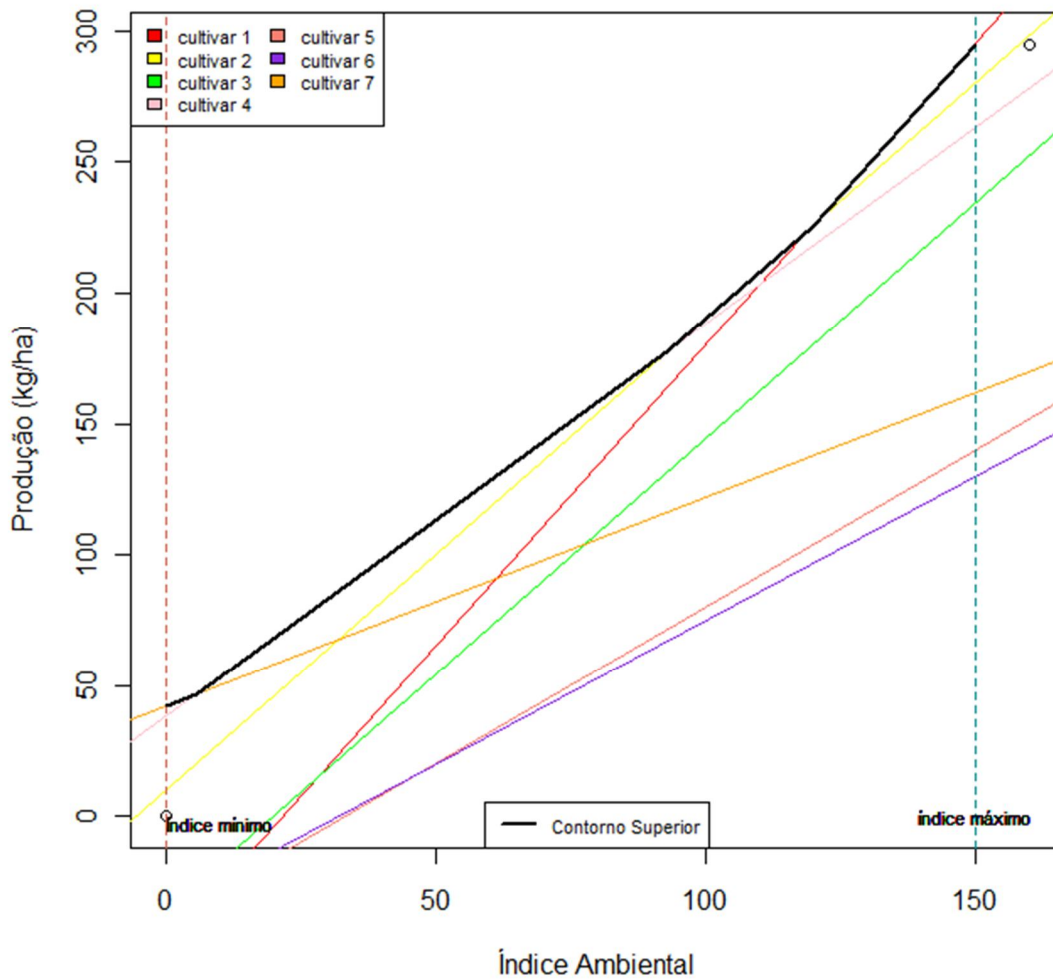


Figura 4.6. Intersecção de várias rectas com declives diferentes.

O caso descrito na alínea b) requer mais cuidado na sua execução pois existem diversas possibilidades que o integram. Por conseguinte vamos estabelecer alguns resultados que caracterizam o traçado do contorno superior

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. Para quaisquer $j \neq 1, j'$ e para $\theta_{1,j} < x < \theta_{1,j'}$, tem-se a desigualdade $\tilde{\alpha}_{j'} + \tilde{\beta}_{j'}x > \tilde{\alpha}_1 + \tilde{\beta}_1x > \tilde{\alpha}_j + \tilde{\beta}_jx$ e portanto o cultivar com índice j' é dominante. ■

Suponhamos agora que em $\theta = \theta_1$ intersectam a recta de índice 1 mais do que uma recta de regressão. Seja F o conjunto dos índices dessas rectas e j' o maior índice do conjunto F . Estabeleçamos a

Proposição 4.3. Se j' é o maior índice para o qual $\theta_1 = \theta_{1,j'}$, o cultivar com índice j' é dominante.

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

Proposição 4.4. Sendo $1 < j < j'$, com j' o maior índice do conjunto F , o cultivar com índice j não é dominante.

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

Na figura seguinte podemos verificar uma situação prevista pela proposição 4.4, onde termos dois cultivares dominantes. (Figura 4.7)

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

```
> dados
```

```
ordenada declive
1      10      1.5
2      30      1.1
3      40      0.9
> dados_ord<-dados[order(-dados$declive,-dados$ordenada),]
> dados_ord

ordenada declive
1      10      1.5
2      30      1.1
3      40      0.9
> uppercontour(dados,0,100)

O cultivar 1 é dominante entre 50 e 100
O cultivar 3 é dominante entre 0 e 50

Os restantes cultivares não fazem parte do contorno superior!
```

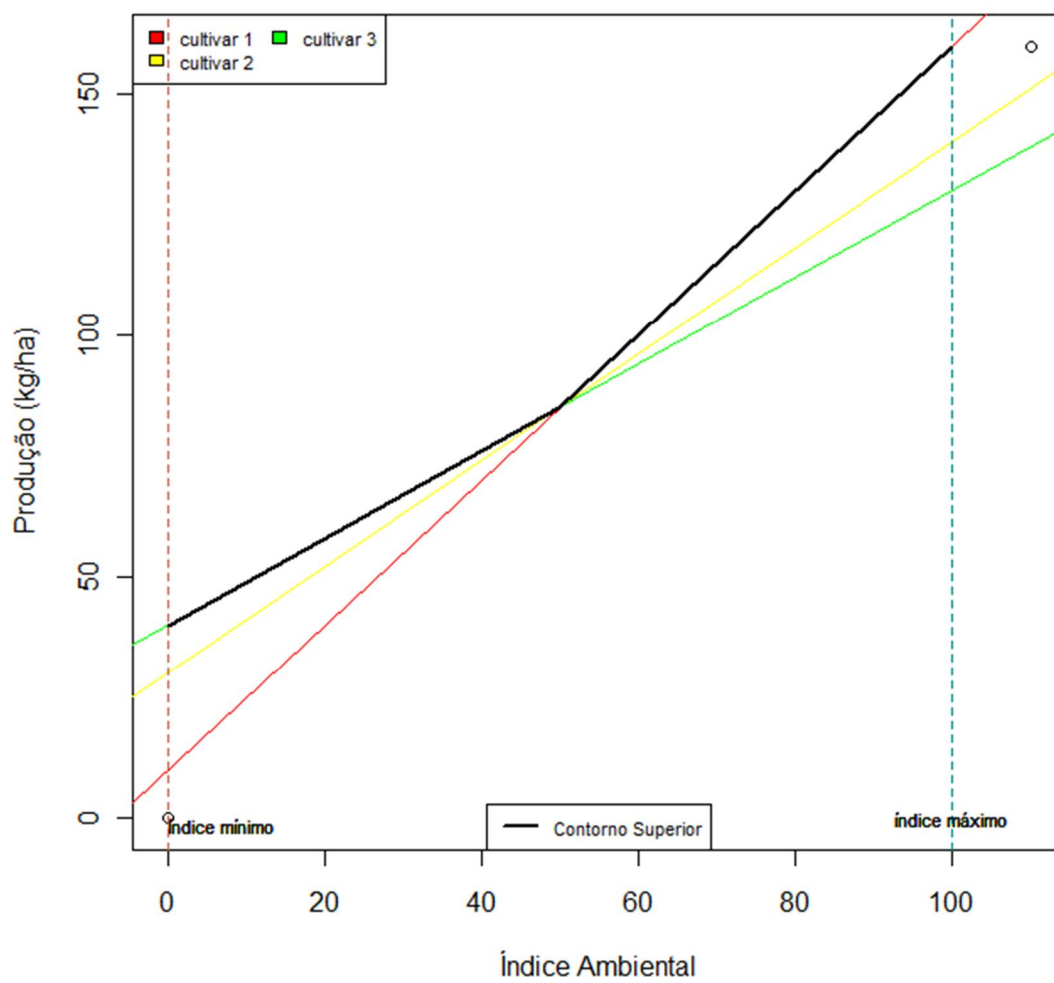


Figura 4.7. Intersecção de três rectas com declives diferentes no mesmo ponto.

4.2. ACR Lineares Completadas

Quando dispomos de cultivares com resposta óptima, o principal objectivo é determinar a forma de seleccionar cultivares com capacidade óptima de resposta em índices ambientais altos. Contudo existe um problema, de acordo com Oliveira (2007) não só se deve atender à produção física como também ao valor da mesma.

Devemos abordar este problema dando importância aos valores de outra variável, sendo que neste caso devemos dar importância ao peso específico, estimando-se coeficientes que traduzam o valor da produção dos vários cultivares. Por conseguinte passa-se a aplicar a Análise Conjunta de Regressões às produções corrigidas dadas pelos produtos dos coeficientes estimados pelas produções físicas.

Interessará averiguar se o método clássico e o método corrigido conduzem à escolha do mesmo cultivar como tendo capacidade óptima de resposta. O método do contorno superior aplicado às regressões ajustadas para as produções física e corrigida, correspondentes aos diferentes cultivares e o método da inclinação máxima que conduz à “selecção” do cultivar com inclinação máxima, aplicando-se este método aos dois conjuntos de regressões.

Oliveira (2007) apresenta o processo detalhado do Método do Contorno Superior quando dispomos cultivares com resposta óptima (ver Oliveira (2007), págs 55 a 60)

Capítulo 5.

Método prático para detecção do Contorno Superior

No presente capítulo, pretendemos inicialmente descrever algoritmicamente o que iremos efectuar no programa para a detecção do contorno superior.

No segundo ponto do mesmo capítulo pretendemos mostrar o fluxograma do mesmo programa, indicando as sucessivas etapas que o programa irá percorrer para finalmente indicar que o Contorno Superior.

Finalmente no terceiro ponto deste mesmo capítulo iremos exemplificar a aplicabilidade do programa, designado por "uppercontour", a dados reais.

5.1. Algoritmo

Para que a detecção do Contorno Superior seja efectuada de forma correcta e eficaz devemos seguir os 8 passos seguintes, onde no final dos mesmos obteremos a indicação gráfico do Contorno Superior.

O mesmo programa irá devolver a indicação de, entre que valores, os cultivares são dominantes.

1º Passo:

Importar a base de dados com declives e ordenadas das rectas de regressão lineares dos cultivares em estudo.

2º Passo:

Ordenar as rectas de regressão segundo os declives (β) das mesmas, de modo descendente.

Caso os declives sejam iguais, devemos ordenar de forma descendente as rectas com os declives iguais segundo a ordenada (α).

3º Passo:

Indicar θ_{min} e θ_{max} que são os limites mínimos e máximos dos índices ambientais ajustados para o presente estudo.

4º Passo:

Se $\beta_j = \beta_{j+1}, j = 1, \dots, J - 1$ e $\alpha_j > \alpha_{j+1}, j = 1, \dots, J - 1$ só a recta j faz parte do contorno superior. A recta $j + 1$ será retirada do estudo.

5º Passo:

Se $\beta_j \neq \beta_{j'}$, temos de determinar os pontos de intersecção em cada conjunto de duas rectas e registar o mesmo valor.

$$\theta_{j,j'} = \frac{\alpha_j - \alpha_{j'}}{\beta_j - \beta_{j'}}, j \neq j'; j, j' = 1, \dots, J$$

6º Passo: Para $j, j=1, \dots, J$

Relativamente ao cultivar com recta de regressão com maior declive (cultivar 1), determinar o ponto máximo $\bar{\theta}_1$ da intersecção da recta do cultivar 1 com as restantes rectas de regressão.

7º Passo:

Determinar a localização no intervalo $[\theta_{min}, \theta_{max}]$ da posição de $\bar{\theta}_1$.

Se:

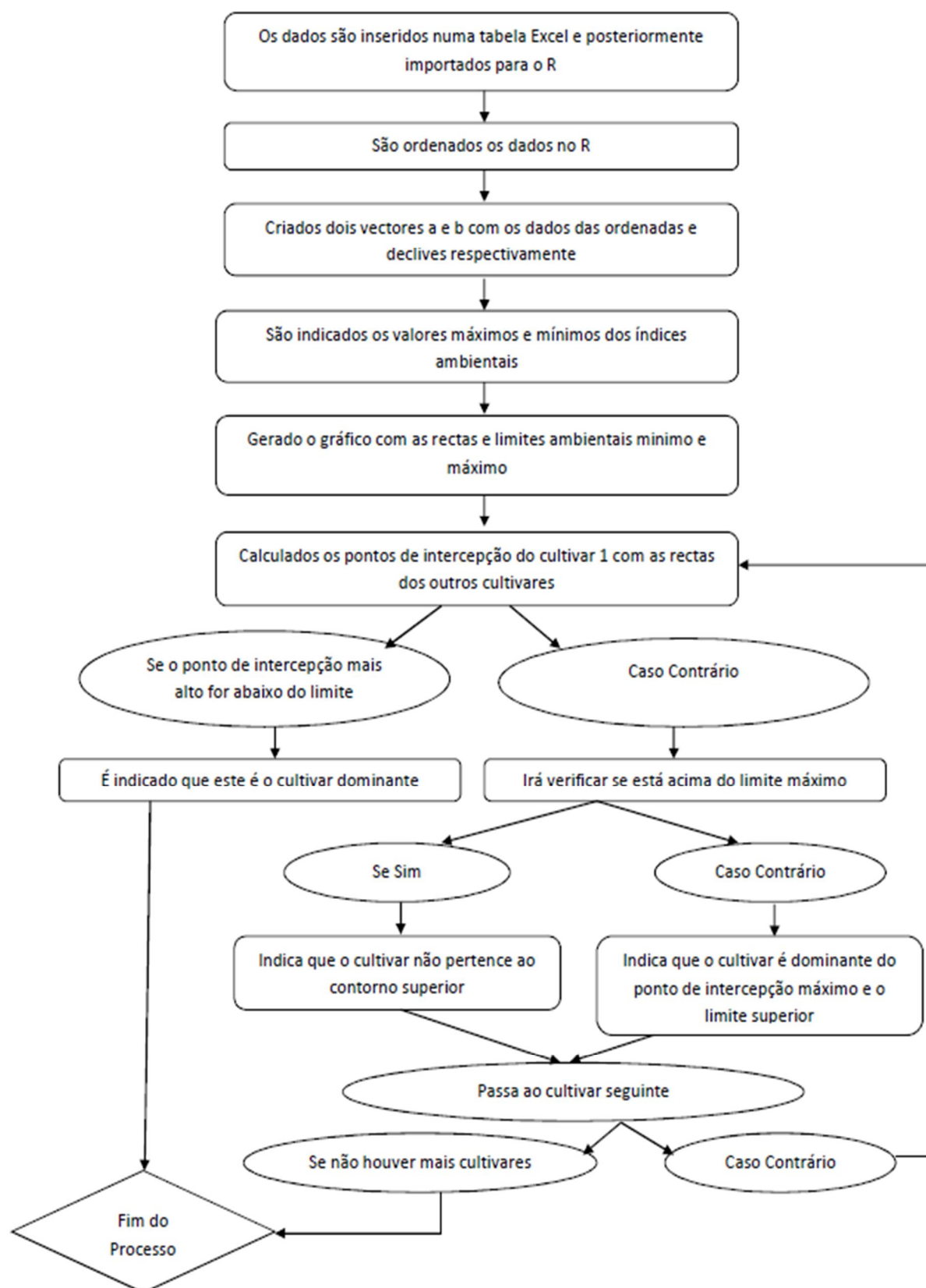
- $\bar{\theta}_1 \leq \theta_{min}$ o cultivar 1 é dominante em todo o processo. Passa-se ao Passo 8.
- $\bar{\theta}_1 \geq \theta_{max}$ o cultivar 1 não é dominante em todo o processo. Volta-se ao Passo 6 com o cultivar seguinte.
- $\theta_{min} < \bar{\theta}_1 < \theta_{max}$ no intervalo $[\bar{\theta}_1; \theta_{max}]$ o cultivar 1 é dominante e no intervalo $[\theta_{min}; \bar{\theta}_1]$ existem outros dominantes. Voltar ao passo 6 com o próximo cultivar.

8º Passo:

Devolver o gráfico com as rectas de regressão ajustadas para os cultivares em estudo e destacado no mesmo gráfico o contorno superior.

Devolver entre que índices ambientais os cultivares são dominantes e quais os cultivares dominantes.

5.2. Fluxograma



5.4. Exemplo de Aplicação

Pretendemos aplicar o mesmo a situações reais. Por conseguinte, utilizando os resultados de Oliveira (2007, página 29) resultantes da aplicação do algoritmo de Zig-Zag. Na Tabela 5.1 encontram-se os coeficientes de regressão ajustados que iremos usar para aplicar o método do contorno superior.

Segundo Oliveira (2007) os mesmos resultados são referentes à produção de aveia, obtidos em ensaios realizados nos anos de 1993 a 1997, pela Estação Nacional de Melhoramento de Plantas.

Tabela 5.1. Coeficientes de Regressão Ajustados

Cultivar	1993		1994		1995		1996		1997	
	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$
AE8901	-108.46	1.270								
AE9004	-137.07	1.205	207.10	0.928						
AE8902	-37.35	1.026								
AE9005	-7.76	1.010	-342.74	1.248						
S.VICENTE	-8.26	0.988								
AE9002	12.96	0.982	-231.79	1.110	67.51	0.972	-595.26	1.255	374.19	0.884
ST	49.60	0.950	196.52	0.886	32.11	1.004	679.48	0.906	220.15	1.128
ALEIXO										
AE9001	59.11	0.909	9.94	0.911						
AE9003	67.56	0.897	13.94	1.005	-79.20	1.235	886.51	0.914	-186.81	1.457
AE8801	91.26	0.890								
AVON	60.58	0.815	90.6	0.928	73.46	0.903	128.66	0.933	-69.67	0.919
AE9303			-122.87	1.082	0.12	1.010	67.49	0.970	-111.66	1.010
AE9101			-85.04	0.969	-249.91	1.210	63.97	0.947	-42.37	0.943
AE9302			170.78	0.948	-24.32	1.010	-197.36	1.117	-14.22	1.103
AE9301			168.33	0.936	130.79	0.910	835.84	0.728	234.15	0.682
AE9401					-72.59	1.052	-1079.90	1.263	-356.53	1.179
AE9402					68.19	0.872	-156.39	0.991	-32.65	0.891
AE9403					73.46	0.713	-268.83	0.824	-40.40	0.891

Neste seguimento, relativamente ao ano de 1993 obtivemos os seguintes resultados

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

```
> dados
```

```
      ordenada declive
1    -108.46    1.270
2    -137.07    1.205
3     -37.35    1.026
4     -7.76    1.010
5     -8.26    0.988
6     12.96    0.982
7     49.60    0.950
8     59.11    0.909
9     67.56    0.897
10    91.26    0.890
11    60.58    0.815
```

```
> dados_ord
```

```
      ordenada declive
1    -108.46    1.270
2    -137.07    1.205
3     -37.35    1.026
4     -7.76    1.010
5     -8.26    0.988
6     12.96    0.982
7     49.60    0.950
8     59.11    0.909
9     67.56    0.897
10    91.26    0.890
11    60.58    0.815
```

```
> uppercontour(dados,0,5000)
```

O cultivar 1 é dominante entre 525.5789 e 5000

O cultivar 10 é dominante entre 0 e 525.5789

Os restantes cultivares não fazem parte do contorno superior!

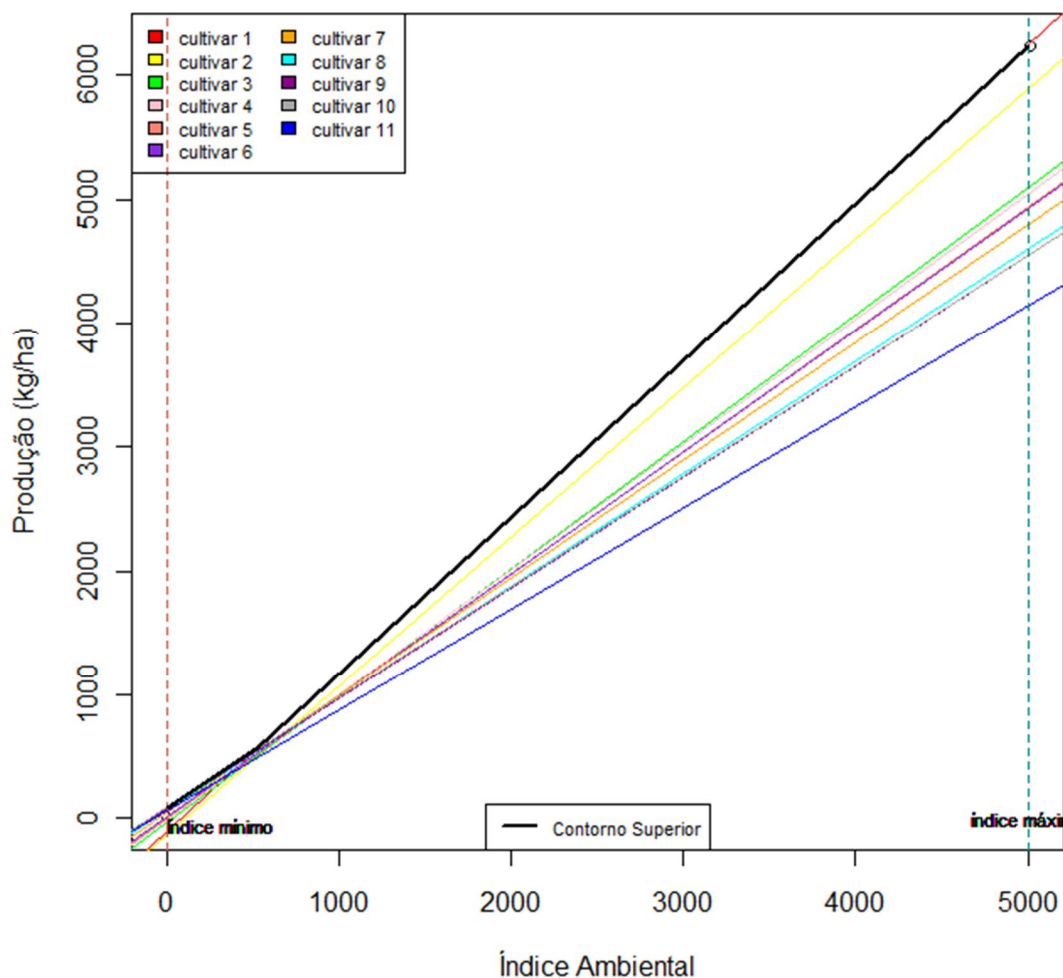


Figura 5.1. Aplicação do Método do Contorno Superior aos dados de 1993 ($J=11$).

Após a ordenação dos cultivares, considerando um índice ambiental mínimo de 0 e um índice ambiental máximo de 5000, verificamos que apenas dois cultivares eram dominantes, sendo os restantes não dominantes, tal como o programa indicou. Esse mesmo facto também pode ser verificado no gráfico resultante do mesmo programa (Figura 5.1).

No que concerne ao ano de 1994 obtivemos os resultados seguintes

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

```
> dados
```

```
      ordenada declive
1      207.10    0.928
2     -342.74    1.248
3     -231.79    1.110
4      196.52    0.886
5         9.94    0.911
6       13.94    1.005
7       90.60    0.928
8     -122.87    1.082
9      -85.04    0.969
10     170.78    0.948
11     168.33    0.936
```

```
> dados_ord
```

```
      ordenada declive
2     -342.74    1.248
3     -231.79    1.110
8     -122.87    1.082
6       13.94    1.005
9      -85.04    0.969
10     170.78    0.948
11     168.33    0.936
1      207.10    0.928
7       90.60    0.928
5         9.94    0.911
4      196.52    0.886
```

```
> uppercontour(dados,0,5000)
```

O cultivar 1 é dominante entre 1718.25 e 5000

O cultivar 8 é dominante entre 0 e 1718.25

Os restantes cultivares não fazem parte do contorno superior!

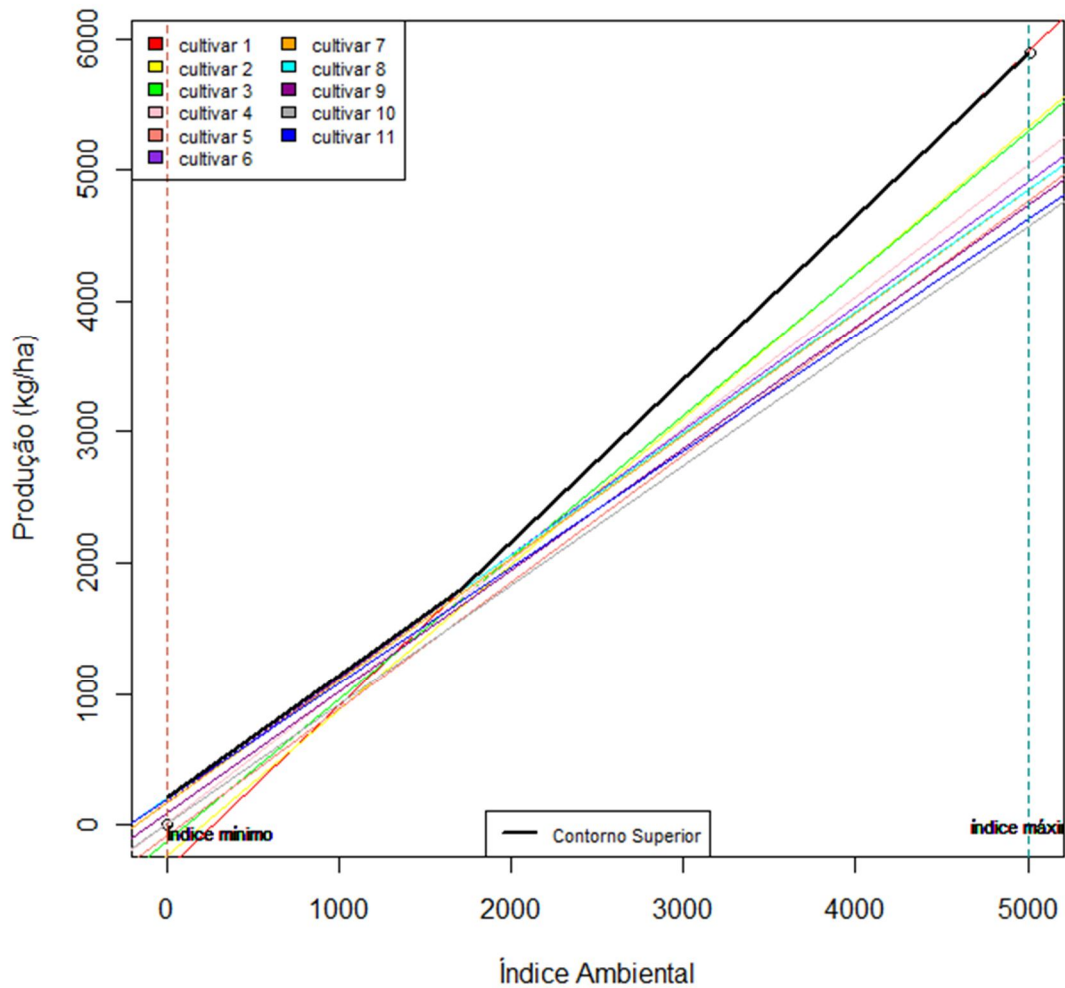


Figura 5.2. Aplicação do Método do Contorno Superior aos dados de 1994 (J=11).

Após a ordenação dos cultivares, considerando um índice ambiental mínimo de 0 e um índice ambiental máximo de 5000, verificamos que apenas dois cultivares são dominantes, sendo os restantes não dominantes, tal como o programa indicou. Esse mesmo facto também pode ser verificado no gráfico resultante do mesmo programa (Figura 5.2).

Relativamente a 1995 os resultados obtidos foram os seguintes

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

```
> dados
```

```
      ordenada declive
1      67.51    0.972
2      32.11    1.004
3     -79.20    1.235
4      73.46    0.903
5       0.12    1.010
6    -249.91    1.210
7     -24.32    1.010
8     130.79    0.910
9     -72.59    1.052
10     68.19    0.872
11     73.46    0.713
```

```
> dados_ord
```

```
      ordenada declive
3     -79.20    1.235
6    -249.91    1.210
9     -72.59    1.052
5       0.12    1.010
7     -24.32    1.010
2      32.11    1.004
1      67.51    0.972
8     130.79    0.910
4      73.46    0.903
10     68.19    0.872
11     73.46    0.713
```

```
> uppercontour(dados,0,5000)
```

O cultivar 1 é dominante entre 646.1231 e 5000

O cultivar 8 é dominante entre 0 e 646.1231

Os restantes cultivares não fazem parte do contorno superior!

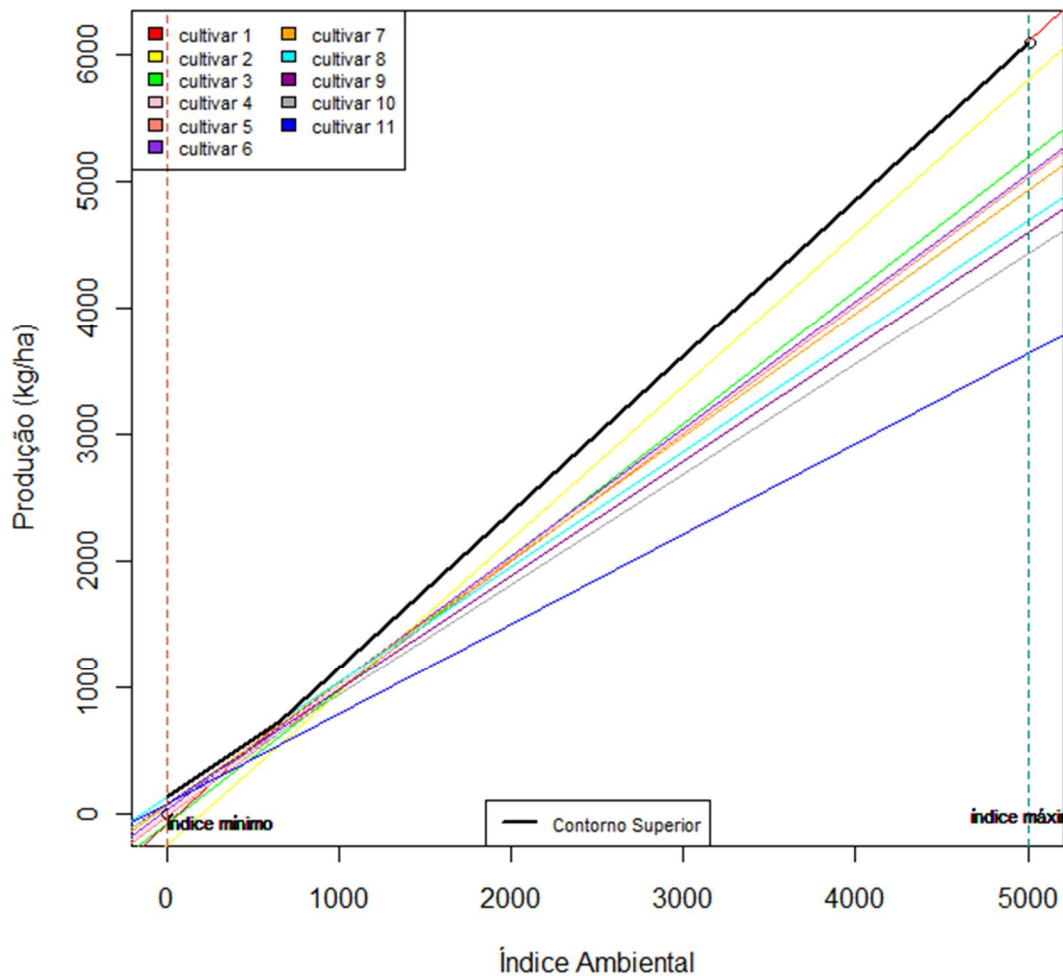


Figura 5.3. Aplicação do Método do Contorno Superior aos dados de 1995 (J=11).

Após a ordenação dos cultivares, considerando um índice ambiental mínimo de 0 e um índice ambiental máximo de 5000, verificamos que apenas dois cultivares são dominantes, sendo os restantes não dominantes, tal como o programa indicou. Esse mesmo facto também pode ser verificado no gráfico resultante do mesmo programa (Figura 5.3).

Em 1996 os resultados são os seguintes

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

> dados

  ordenada declive
1  -595.26   1.255
2   679.48   0.906
3   886.51   0.914
4   128.66   0.933
5    67.49   0.970
6    63.49   0.947
7  -197.36   1.117
8   835.84   0.728
9 -1079.90   1.263
10 -156.39   0.991
11 -268.83   0.824

> dados_ord

  ordenada declive
9 -1079.90   1.263
1  -595.26   1.255
7  -197.36   1.117
10 -156.39   0.991
5    67.49   0.970
6    63.49   0.947
4   128.66   0.933
3   886.51   0.914
2   679.48   0.906
11 -268.83   0.824
8   835.84   0.728

> uppercontour(dados,0,5000)

[1] "O cultivar 1 não faz parte de contorno superior"
O cultivar 2 é dominante entre 4345.367 e 5000
O cultivar 8 é dominante entre 0 e 4345.367
```

Os restantes cultivares não fazem parte do contorno superior!

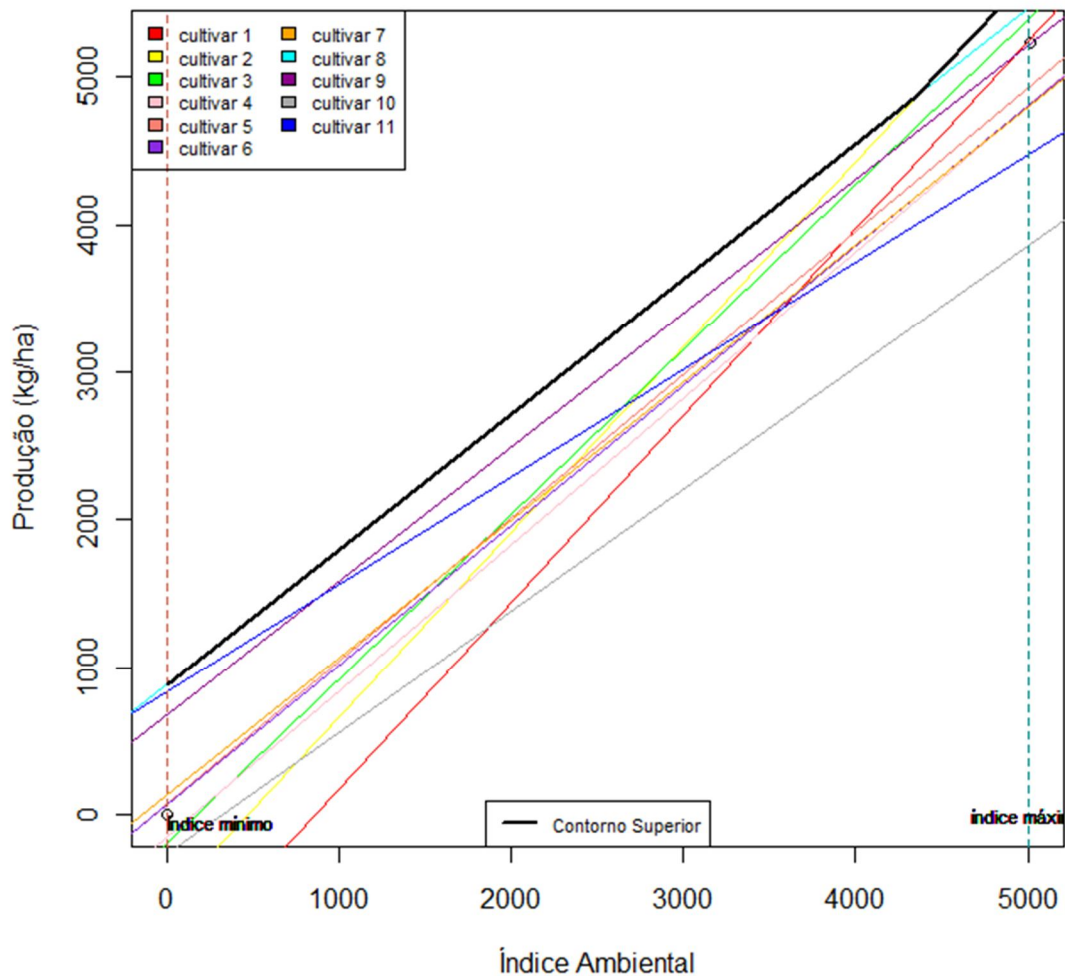


Figura 5.4. Aplicação do Método do Contorno Superior aos dados de 1996 (J=11).

Após a ordenação dos cultivares, considerando um índice ambiental mínimo de 0 e um índice ambiental máximo de 5000, verificamos que três cultivares eram dominantes, sendo os restantes não dominantes, tal como o programa indicou. Esse mesmo facto também pode ser verificado no gráfico resultante do mesmo programa (Figura 5.4).

Por fim, os resultados referentes ao ano de 1997 são os seguintes

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

```
> dados
```

```
      ordenada declive
1      374.19    0.884
2      220.15    1.128
3     -186.81    1.457
4      -69.67    0.919
5     -111.66    1.010
6      -42.37    0.943
7      -14.22    1.103
8      234.15    0.682
9     -356.53    1.179
10     -32.65    0.891
11     -40.40    0.891
```

```
> dados_ord
```

```
      ordenada declive
3     -186.81    1.457
9     -356.53    1.179
2      220.15    1.128
7      -14.22    1.103
5     -111.66    1.010
6      -42.37    0.943
4      -69.67    0.919
10     -32.65    0.891
11     -40.40    0.891
1      374.19    0.884
8      234.15    0.682
```

```
> uppercontour(dados,0,5000)
```

O cultivar 1 é dominante entre 1236.960 e 5000

O cultivar 3 é dominante entre 631.3115 e 1236.960

O cultivar 10 é dominante entre 0 e 631.3115

Os restantes cultivares não fazem parte do contorno superior!

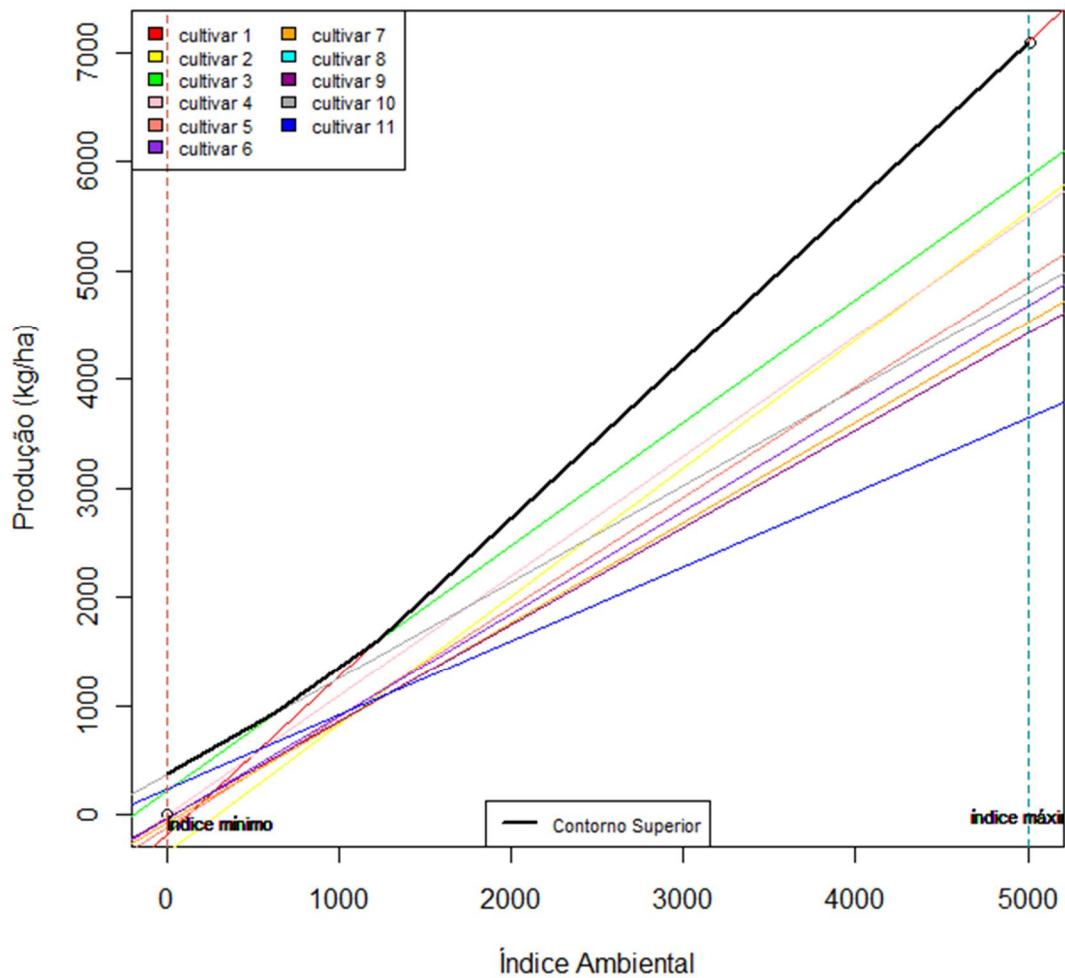


Figura 5.5. Aplicação do Método do Contorno Superior aos dados de 1997 ($J=11$).

Após a ordenação dos cultivares, considerando um índice ambiental mínimo de 0 e um índice ambiental máximo de 5000, verificamos que em 1997 também três cultivares eram dominantes, sendo os restantes não dominantes, tal como o programa indicou. Esse mesmo facto também pode ser verificado no gráfico resultante do mesmo programa (Figura 5.5).

Conclusões e trabalho futuro

Neste trabalho foram apresentados os principais resultados da técnica ACR para comparação e selecção de cultivares, onde destacámos o método do contorno superior.

Este estudo permitiu o desenvolvimento dum método expedito para a determinação do Contorno Superior, usando a linguagem R ficando demonstrada a versatilidade e aplicabilidade desta linguagem que é do tipo *OpenSource*.

Através do programa criado, , é possível determinar o Contorno Superior. Deste modo podemos obter representações gráficas do Contorno e a indicação dos cultivares, sendo os mesmos distinguidos por cores. Devemos para isso, apenas indicar onde se encontra a base de dados com os coeficientes de regressão ajustados. Seguidamente usando a função "uppercontour" descrita do modo seguinte

```
> uppercontour(dados,"índice ambiental mínimo","índice ambiental máximo")
```

É possível obter os resultados pretendidos.

Como devemos ter numa base de dados previamente construída, os coeficientes de regressão ajustados pelo algoritmo de Zig-Zag, o mesmo facto provoca algum trabalho extra aos investigadores. Deste modo propomo-nos no futuro desenvolver outras funções em R, nomeadamente para a implementar o algoritmo Zig-Zag e outros procedimentos que possam contribuir para a criação de um *package* possível de ser usado de forma simples e célere na comparação e selecção de cultivares.

Anexo

Anexo A

Programa em R

```
> uppercontour<-function(dados,x1,x2){
+
+ dados_ord<-dados[order(-dados$declive,-dados$ordenada),]
+
+ a<-c(dados_ord$ordenada) #cria um vector com as ordenadas
+
+ b<-c(dados_ord$declive) #cria um vector com os declives
+
+ n<-length(dados_ord$declive) #atribui a "n" o número de
declives
+
+ x<-c(x1,x2) #atribui a "x" os índices
ambientais mínimo(x1) e máximo(x2)
+
+ inter<-0
+
+ z<-c("red","yellow","green","pink","black","blueviolet",
+"orange","cyan","darkmagenta","darkgrey","blue","brown",
+"azure3","chocolate3","darkolivegreen1","darkorchid1",
+"cadetblue4","darkred","chartreuse2","darkgoldenrod3",
+"burlywood","darkcyan") #"z" é um vector com 22 cores
+
+ Rendimento<-c(0,x[2]*b[1]+a[1]) #eixo do "y" no gráfico
bidimensional
```

```
+  
  
+ Indice_Ambiental<-c(0,x[2]+10) #eixo do "x" no gráfico  
bidimensional  
  
+  
  
+ plot(Indice_Ambiental,Rendimento,xlab="Índice  
Ambiental",ylab="Produção (kg/ha)") #gráfico bidimensional  
  
+  
  
+ #o ciclo abaixo permite a criação do gráfico com os  
diversos cultivares devidamente legendados  
  
+ for (i in 1:n){  
  
+  
  
+ abline(a[i],b[i],col=z[i]) #insere no gráfico as "n"  
rectas, cada uma de cor diferente  
  
+  
  
+ abline(v=x[1],col="coral3",lty=2) #insere uma linha  
vertical (índice ambiental mínimo)de uma determinada cor e  
a tracejado.  
  
+  
  
+ text(x[1],0,"índice mínimo",cex=0.7,adj=c(0,1)) #atribui  
à linha vertical x[1] o nome "índice mínimo"  
  
+  
  
+ abline(v=x[2],col="cyan4",lty=2) #insere uma linha  
vertical (índice ambiental máximo)de uma determinada cor e  
a tracejado.  
  
+  
  
+ text(x[2],0,"índice máximo",cex=0.7)} #atribui à linha  
vertical x[2] o nome "índice máximo"  
  
+  
  
legend("topleft",paste("cultivar",1:n),cex=0.7,ncol=2,fill=  
z) #introduz uma legenda no canto superior esquerdo com o  
nome dos cultivares e cor correspondente  
  
+
```

```
+ #no ciclo abaixo o programa irá comparar o cultivar 1 com
os restantes
+ for (i in 1:n){
+
+ inter[i]<-c((a[1]-a[i])/(b[i]-b[1]))} #calcula todas as
intercepções entre o cultivar 1 e os restantes cultivares
+
+ inter[is.infinite(inter)]<-0 #dentro do vector inter
todas intercepções infinitas tomam o valor zero
+
+ inter[is.na(inter)]<-0 #dentro do vector inter todas
intercepções indeterminadas tomam o valor zero
+
+ interception<-max(na.exclude(inter))          #a variável
"interception" assume o máximo de todas as intersecções
+
+ #o programa irá comparar a variável "interception" com os
índices ambientais mínimo e máximo e toma a decisão de
acordo com as indicações do ciclo "if"
+ if (x[1]>interception) {print('O cultivar 1 é dominante')}
+
+ t<-1
+
+ y<-a[1]+b[1]*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
+
+ x[2]=x[1]-1} else
+
```

```
+ if (x[2]<=interception) {print('O cultivar 1 não faz
parte de contorno superior')}
+
+ t<-0} else
+
+ {cat ('\n O cultivar 1 é dominante entre', interception,
'e', x[2])
+
+ m<-c(interception,x[2])
+
+ y<-a[1]+b[1]*m
+
+ lines(m,y,type="l",lwd=2)          #o programa irá escrever
colocar no gráfico o contorno superior destacado entre o
valor da variável "interception" e o índice ambiental
máximo
+
+ x[2]<-interception          #o índice máximo passa a tomar o
valor da variável "interception"
+
+ t<-1}
+
+ w<-1
+ #no ciclo "for" abaixo o programa compara os valores de
intercepção no vector "inter", caso dois ou mais valores
forem iguais o proximo cultivar a ser selecionado é o que
tiver maior rendimento antes da intersecção
+ for(i in 1:n)
+
+ {if (inter[w]<inter[i])
+
+ }
```

```
+ w<-i else if(inter[w]==inter[i] & a[w]+b[w]*(inter[w]-
0.05)<a[i]+b[i]*(inter[i]-0.05)) w<-i else
+
+ w<-w}
+
+ #o ciclo "while" irá efectuar uma comparação entre
cultivares até que o índice máximo seja menor que o índice
mínimo
+ while(x[2]>x[1])
+
+ {for (j in 1:n)
+
+ inter[j]<-c((a[w]-a[j])/(b[j]-b[w]))
+
+ inter[is.infinite(inter)]<-0
+
+ inter[is.na(inter)]<-0
+
+ for (j in 1:n) if (inter[j]>=x[2]) inter[j]<-0
+
+ interception<-max(na.exclude(inter))
+
+ if (x[1]>interception) {cat('\n O cultivar', w, 'é
dominante entre', x[1], 'e', x[2])
+
+ y<-a[w]+b[w]*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]<interception) {cat('\n O cultivar',w,' não faz
parte de contorno superior')
+
+ t<-t} else
+
+ {cat ('\n O cultivar',w,'é dominante entre',interception,
'e', x[2])
+
+ m<-c(interception,x[2])
+
+ y<-a[w]+b[w]*m
+
+ lines(m,y,type="l",lwd=2)          #o programa irá escrever
colocar no gráfico o contorno superior destacado entre o
valor da variável "interception" e o índice ambiental
máximo
+
+ t<-t+1}
+
+ x[2]<-interception #o índice máximo passa a tomar o valor
da variável "interception"
+
+ for(j in 1:n)
+
+ {if (inter[w]<inter[j])
+
+ }
```

```
+ w<-j else if(inter[w]==inter[i] & a[w]+b[w]*(inter[w]-
0.05)<a[j]+b[j]*(inter[j]-0.05)) w<-j else
+
+ w<-w}}
+
+ if (t<n) cat('\n Os restantes cultivares não fazem parte
do contorno superior!') #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
```

Bibliografia

- [1] **Adler, Joseph (2009)** – R in a Nutshell: A desktop quick reference. O'Reilly. Sebastopol, CA.
- [2] **Bose, R. C. (1950)** - Least Squares aspects of analysis of variance. Mimeo Series 9, Institute of Statistics, University of North Carolina, Chapel Hill.
- [3] **Crawley, Michael J. (2007)** - The R book, 950p. Wiley.
- [4] **Eberhart, S. A. and Russel, W. A. (1966)** - Stability parameters for comparing varieties - Crop. Sci, 6 : 36 - 40.
- [5] **Ferreira, T.; Mexia, J. T.; Baeta, J.; Maças, B.; Gusmão, L; Bagulho, F. & 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. In “V Reunião Portuguesa sobre Triticales” – Elvas, Maio 1990 – Melhoramento, vol. 32.
- [6] **Faraway, Julian J. (2002)** - Practical Regression and Anova using R, 213p. Disponível em: <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>. Acedido a 21 de Novembro de 2009.
- [7] **Faraway, Julian J. (2004)** - Linear Models with R, 240p., Chapman & Hall/CRC, EUA.
- [8] **Faraway, Julian J. (2006)** - Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, 324p., Chapman & Hall/CRC, EUA.
- [9] **Farnsworth, Grant V. (2008)** – Econometrics in R. Disponível em: <http://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf>. Acedido a 21 de Setembro de 2010.
- [10] **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

-
- [11] **Finlay, K. W. and Wilkison, G. N. (1963)** - The analysis of adaptation in a plant breeding programme. - Aust. J.Aric. Res., 14 : 742 - 754.
- [12] **Fox, John (2002)** - An R and S-PLUS Companion to Applied Regression, 312p., Sage Publications, Inc. USA.
- [13] **Giraldes et al. (1995)** – Curso de Algebra Linear e Geometria Analítica. Mc Graw-Hill.
- [14] **Girke, Thomas [s.d]** – Programming in R. Disponível em: <http://manuals.bioinformatics.ucr.edu/home/programming-in-r>. Acedido a 20 de Janeiro de 2010.
- [15] **Gusmão, L. (1985)** – An adequate design for regression analysis of yield trials. Theor. Appl. Genet., 71: 314 - 319.
- [16] **Gusmão, L. (1986a)** – Inadequacy of blocking in cultivar yield trials. Theor. Appl. Genet., 72: 98 - 104.
- [17] **Gusmão, L. (1986b)** – A interacção genótipo x ambiente e a comparação de cultivares de cereais. Tese de doutoramento. Instituto Superior de Agronomia. Universidade Técnica de Lisboa, Lisboa.
- [18] **Gusmão, L. (1990)** - Avaliação de cultivares (uma perspectiva experimental diferente). Revista de Ciências Agrárias - n_2, vol. 13: 11 - 18.
- [19] **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.
- [20] **Gusmão, L. ; Mexia, J. T. and Baeta, J. (1992a)** – Trimmed Joint Regression: a new approach to the Joint Regression Analysis for cultivar relative performance evaluation. Theoretical and Applied Genetics, 84: 735 – 738.
- [21] **Gusmão, L.; Baeta, J.; Mexia J. T.; Bagulho, F.; Maças, B. & Coutinho, J. (1992b)** – Statistical analysis of randomised complete block designs for Barley cultivar yield trials. Rachis, 11 : 14 – 18.

- [22] **Gusmão, L.; Rebelo, D. C. ; Miranda, M. J.; Baeta, J. and Vaz, F. (1993a)** – Assessment of specific genotype x environment interactions in *Lupinus luteus* L.. *Brotéria Genética*, Lisboa, 14 : 141 - 144.
- [23] **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 speci_c interactions. ABST. - VII International Lupin Conference, Évora, 18 a 23 de Outubro, Theme:3 , n_ 32.
- [24] **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.
- [25] **Johnson, Paul (2001)** - Rtips. Site: <http://pj.freefaculty.org/R/Rtips.html>, Acedido a 21 de Novembro de 2009.
- [26] **Júnior, Paulo Justiniano Ribeiro (2005)** - Curso sobre o programa computacional R. Disponível em: <http://www.leg.ufpr.br/Rpira/Rpira.pdf>. Acedido a 25 de Maio de 2010.
- [27] **Kageyama, S. (1984)** - Connected designs with the minimum number of experiments units. *Linear Statistical Inference, Lecture Notes in Statistics* n. 35, Springer Verlag, New york : 99 - 117.
- [28] **Kastenmeier, Rainer, 2008**, *Joint Regression Analysis of Insurance Claims and Claim Sizes*, 106p., Technische Universität München, Alemanha.
- [29] **Lin, C. S., Binns, M. R. and Lefkovich, L. P. (1986)** - Stability analysis: Where Do We Stand? *Crop. Sci.*, 26: 894 - 900.
- [30] **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 Triticales” – Elvas, Maio 1990 – Melhoramento, vol. 32.

- [31] **Mexia, J. T.; Gusmão, L. e Baeta, J. (1991)** - Analysis of cultivar yield trials designed in Randomize Complete Blocks. - Revista de Ciências Agrárias, 14: 59 - 63.
- [32] **Mexia, J. T. (1995)** - Introdução à Inferência Estatística Linear. Edições Lusófonas, Lisboa.
- [33] **Mexia J. T.; Amaro, A. P., Gusmão, L. and Baeta J. (1997)** – Upper contour of a Joint Regression Analysis, J. Genet & Breed., 51 : 253 - 255.
- [34] **Mexia J. T. ; Pereira, D. G. and Baeta, J. (1999)** – L_2 environmental indexes. Biometrical Letters, 36 : 137 – 143.
- [35] **Mooers, C. A. (1921)** - The agronomic placement of varieties. - J. Amer. Soc. Aron., 13 : 337 - 352.
- [36] **Ng, M. P. and Grunwald, G. K. (1997)** – Nonlinear regression analysis of the joint-regression model. Biometrics, 53 : 1366 – 1372.
- [37] **Oliveira, Amílcar Manuel do Rosário (2007)** - Estabilidade em Análise Conjunta de Regressões e Condução Dinâmica de Planos de Melhoramento, pp. III, 1, 119p. Universidade Aberta, Lisboa. Tese de Doutoramento.
- [38] **Oliveira, Amílcar Manuel do Rosário (2008)** - Análise Conjunta de Regressões na Condução de Planos de Melhoramento. Universidade Aberta, Lisboa
- [39] **Patterson, H.D. and Williams, E.R. (1976)** - A new class of resolvable incomplete block designs. Biometrika, 63: 83 - 92.
- [40] **Perkins, J.M.; Jinks, J.L. (1968)** – Environmental and genotype-environmental components of variability. III Multiple lines and crosses. Heredity, 23: 339 – 356.
- [41] **Pereira, D.G; Mexia, J.T. (2003)** – The use of Joint Regression Analysis with special reference to variety selection. Scientific Papers of the Agricultural University of Poznan, Agriculture, 3: 67-74.
- [42] **Pereira, D. G. (2004)** – Análise Conjunta Pesada de Regressões em Redes de Ensaio – Tese de Doutoramento. Universidade de Évora.

- [43] **Pinto, Iola Maria Silvério (2006)** - Análise Conjunta de Regressões e Planos de Melhoramento, pp. V, 5, 129p., FCT – Universidade Nova de Lisboa, Lisboa.
- [44] **Pollock, D. S. G: (1979)** – The Algebra of Econometrics. John Wiley & Sons, Inc.
- [45] **Raghavarao, D. (1971)** - Construction and combinatorial Problems in Design of Experiments. Wiley, New York. Reprinted (1988) by Dover with some addendum.
- [46] **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.
- [47] **Seber, G.A.F. (1980)** - The Linear Hypothesis : A General theory., 2nd ed., Charles Griffin and Co. Londres.
- [48] **Seber, G.A.F. and Lee, J. A. (2003)** - Linear Regression Analysis, 2nd ed., John Willey and Sons - New York.
- [49] **Sheather, Simon J. (2009)** - A Modern Approach to Regression with R, 393p., Springer.
- [50] **Silva, Bruno Fontana; Diniz, Jean e Bortoluzzi, Matias Américo (2009)** – Minicurso de Estatística Básica: Introdução ao Software R. Universidade Federal de Santa Maria. Disponível em: <http://petee01.ctlab.ufsm.br/petee/sites/default/files/Curso%20de%20R%20-%20Apostila%20PET-EE.pdf>. Acedido a 4 de Abril de 2010.
- [51] **Silva, M. F (2007)** - Apresentando o R. Disponível em: <http://marcosfs2006.googlepages.com/apresentandoor>. Acedido a 9 de Outubro de 2010.
- [52] **Torgo, Luís (2009)** - A linguagem R, Programação para a análise de dados, Escolar Editora, Lisboa.
- [53] **Torgo, Luís (2006)** - Introdução à Programação em R, 99p., Faculdade de Economia, Universidade do Porto, Porto.

- [54] **Vance, Ashlee (2009)** - Poder da linguagem R fascina analistas. Site: <http://tecnologia.terra.com.br/interna/0,,OI3439506-EI4801,00-Poder+da+linguagem+R+fascina+analistas.html>. Acedido a 10 de Novembro de 2009.
- [55] **Venables, W. N.; Smith, D. M. et al. (2010)** – An Introduction to R. Disponível em: <http://cran.r-project.org/doc/manuals/R-intro.pdf>. Acedido a 20 de Janeiro de 2010.
- [56] **Westcott, B. (1986)** - Some methods of analyzing genotype-environment interaction *Heredity*, 56: 243 - 253.
- [57] **Williams, E. R., Matheson, A. C., Harwood, C. E. (2002)** - Experimental design and analysis for tree improvement, 2^oEd., 220p., Csiro Publishing, Australia. Disponível em: http://books.google.pt/books?id=pPCX-o0eMMgC&pg=PA84&lpg=PA84&dq=%22joint+regression+analysis%22&source=bl&ots=mqiArngQo&sig=Llh80tz28yQj1s7StPdhFMM-A5Y&hl=pt-PT&ei=X1cBS8fTDsGd4QaB3qjxCw&sa=X&oi=book_result&ct=result&resnum=6&ved=0CCgQ6AEwBQ#v=onepage&q=%22joint%20regression%20analysis%22&f=false. Acedido a 14 de Novembro de 2009.
- [58] **Yates, F. and Cochran, W. G. (1938)** – The analysis of groups experiments. *J. Agric. Sci., Cambridge*, 28 : 556 - 580.
- [59] **[s.n], [s.d]** - An Introduction to R. Site: <http://cran.r-project.org/doc/manuals/R-intro.html>, Acedido a 10 de Novembro de 2009.
- [60] **[s.n], [s.d]** - The R Project for Statistical Computing. Disponível em: <http://www.r-project.org>, Acedido a 10 de Novembro de 2009.
- [61] **[s.n], (2009)** - R Language Definition, 60p. Disponível em: <http://cran.r-project.org/doc/manuals/R-lang.pdf>. Acedido a 20 de Novembro de 2009.
- [62] **[s.n], (2010)** - Oracle dá suporte à linguagem R. Disponível em: <http://www.computerworld.com.pt/2010/06/01/oracle-da-suporte-a-linguagem-de-codigo-aberto-r>. Acedido a 20 de Setembro de 2010.

- [63] **[s.n], [s.d]** - Index of /R-manual/R-devel/library/stats/html. Disponível em: <http://stat.ethz.ch/R-manual/R-devel/library/stats/html/>. Acedido a 30 de Setembro de 2010.