

LAÍS MAYARA AZEVEDO BARROSO

**REGRESSÃO QUANTÍLICA NA AVALIAÇÃO DA  
ADAPTABILIDADE E ESTABILIDADE  
FENOTÍPICA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2014

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

B277r  
2014

Barroso, Laís Mayara Azevedo, 1989-  
Regressão quantílica na avaliação da adaptabilidade e  
estabilidade fenotípica / Laís Mayara Azevedo Barroso. –  
Viçosa, MG, 2014.  
xi, 70f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Moysés Nascimento.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Alfafa - Melhoramento genético. 2. Interação genótipo x  
ambientes. I. Universidade Federal de Viçosa. Departamento de  
Estatística. Programa de Pós-graduação em Estatística Aplicada  
e Biometria. II. Título.

CDD 22. ed. 633.31

LAÍS MAYARA AZEVEDO BARROSO

**Regressão quantílica na avaliação da adaptabilidade e estabilidade fenotípica**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 17 de fevereiro de 2014

---

Fabyano Fonseca e Silva

---

Ana Carolina Campana Nascimento  
(Coorientadora)

---

Moisés Nascimento  
(Orientador)

“Seria mais fácil fazer como todo mundo faz  
o caminho mais curto, produto que rende mais  
seria mais fácil fazer como todo mundo faz  
um tiro certo, modelo que vende mais...

Mas nós vibramos em outra frequência  
sabemos que não é bem assim  
se fosse fácil achar o caminho das pedras  
tantas pedras no caminho não seria ruim.”

(Engenheiros do Hawaii)

“Nunca deixe que lhe digam que não vale a pena  
Acreditar no sonho que se tem”

(Renato Russo)

## AGRADECIMENTOS

Agradeço a Deus por ter me dado força para realização de mais este sonho, por ouvir minhas preces e por ter colocado pessoas maravilhosas no meu caminho que muito me auxiliaram nesta etapa que chega ao fim.

Aos meus pais, Adelson e Inêz, por sonharem junto comigo e por serem minha base. Tudo o que sou hoje agradeço a vocês. Obrigada pelas orações, incentivos e apoio em todos os momentos. Vocês são meus heróis e exemplos.

Aos meus irmãos Thiago, Livia e Maria Isabel por estarem sempre ao meu lado me apoiando e me fazendo dar muitas risadas. Por todos os momentos que vivemos que só nos ajudaram a crescer e aumentar ainda mais a nossa união.

Ao meu orientador e amigo Moysés Nascimento pelos conselhos, incentivo e confiança depositada na execução deste trabalho. Agradeço pela disponibilidade, atenção e amizade adquirida ao longo destes anos de convivência. Por me incentivar a cada dia mais e por ser um exemplo de profissional a ser seguido.

À professora e co-orientadora Ana Carolina Campana Nascimento pelas sugestões, incentivo e carinho.

Aos professores e co-orientadores Cosme Damião Cruz e Leonardo Lopes Bhering pelas sugestões e pela ajuda.

Ao professor Fabyano Fonseca e Silva por abrir as portas da estatística para mim na Iniciação Científica fazendo com que eu tomasse gosto pela pesquisa, por me apresentar o professor Moysés nesta época para que pudéssemos trabalhar juntos e por ter aceitado participar desta banca.

Aos amigos de mestrado, Édimo, Eliângela, Fanni, Pâmela, Nayara, Lucas e Regiane pelos ótimos momentos e pelas trocas de experiência. A Camila, minha eterna veterana, pelas conversas e pelos conselhos. A Gabi, que ao longo destes anos se tornou uma irmã, por todos os momentos inesquecíveis, pela força e pela amizade. Sem vocês nada disso seria possível.

Aos amigos de Viçosa, em especial a Camila, Matheus, Vinicius, Rafael, Thiago, Victor, Michele, Izabela e Susana por sempre terem uma palavra de incentivo e apoiarem minhas decisões.

Às amigas de Bocaiúva, Maria Eugênia, Marcela, Bárbara, Raíssa, Verônica, Izabella pela valiosa amizade e por torcerem sempre por mim.

À República Sereníssimas, Ana Flávia, Ana Marisa, Tássia, Carol, Júlia, Priscila, Aline, Kéllen e Cristiana, por serem minha família em Viçosa, por tornarem essa caminhada mais tranquila e divertida. Serei eternamente grata a vocês.

Aos meus familiares pelo apoio.

Aos professores e funcionárias do Departamento de Estatística da UFV, pela competência profissional e por todo apoio dado ao longo das minhas atividades acadêmicas.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Estatística Aplicada e Biometria pela oportunidade.

À CAPES, pela concessão da bolsa de estudos.

À todos que de uma maneira ou outra auxiliaram na concretização deste trabalho.

## **BIOGRAFIA**

LAÍS MAYARA AZEVEDO BARROSO, filha de Maria Inez Azevedo Barroso e Antônio Adelson Barroso, nasceu em Bocaiúva, Minas Gerais, em 16 de março de 1989.

Em março de 2007, ingressou no curso de Licenciatura em Matemática na Universidade Federal de Viçosa, Viçosa - MG, graduando-se em janeiro de 2012.

Em março do mesmo ano, iniciou o curso de Mestrado no Programa de Pós-Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, submetendo-se à defesa da dissertação em 17 de fevereiro de 2014.

## SUMÁRIO

RESUMO.....	viii
ABSTRACT.....	x
1 INTRODUÇÃO GERAL.....	1
2 REVISÃO DE LITERATURA.....	4
2.1 Interação Genótipo x Ambiente.....	4
2.2 Metodologia de Adaptabilidade e Estabilidade baseada em Regressão Linear Simples.....	5
2.2.1 Eberhart e Russell (1966).....	5
2.3 Metodologia de Adaptabilidade e Estabilidade baseada em análises não paramétrica.....	7
2.3.1 Adaptabilidade e Estabilidade via regressão não paramétrica.....	8
2.4 Regressão Quantílica.....	9
2.4.1 Introdução.....	9
2.4.2 Quantil como solução para um problema de minimização.....	11
2.4.3 Modelo da RQ.....	13
2.4.4 Estimação.....	14
2.4.5 Intervalo de Confiança.....	18
2.4.5.1 Metodologia baseada em <i>bootstrap</i> .....	18
2.4.6 Qualidade de ajuste na RQ.....	18
2.4.7 Interpretação das estimativas dos parâmetros.....	19
2.4.7.1 RLS.....	19
2.4.7.2 RQ.....	19
REFERÊNCIAS BIBLIOGRÁFICAS.....	21
CAPÍTULO 1.....	25
Adaptabilidade e estabilidade para genótipos de alfafa via regressão quantílica.....	25
RESUMO.....	25
1 Introdução.....	26
2 Material e Métodos.....	27
2.1 RQ em estudos de adaptabilidade e estabilidade.....	27
2.2 Dados simulados.....	30

2.3 Comparação de métodos de adaptabilidade e estabilidade.....	31
2.4 Dados Reais.....	32
2.5 Aspectos Computacionais.....	33
3 Resultados e Discussão.....	34
3.1 Dados Simulados .....	34
3.2 Dados Reais.....	37
4 Conclusões.....	44
REFERÊNCIAS BIBLIOGRÁFICAS.....	45
CONSIDERAÇÕES FINAIS.....	48
APÊNDICE A - Rotinas Computacionais implementadas .....	49
APÊNDICE B – Tabela com os demais genótipos.....	65
APÊNDICE C - Demonstração de que a mediana minimiza a média da distancia absoluta.....	69

## RESUMO

BARROSO, Laís Mayara Azevedo, M.Sc., Universidade Federal de Viçosa, fevereiro de 2014. **Regressão Quantílica na avaliação da adaptabilidade e estabilidade fenotípica.** Orientador: Moysés Nascimento. Coorientadores: Ana Carolina Campana Nascimento, Cosme Damião Cruz e Leonardo Lopes Bhering.

No melhoramento genético de plantas, quando o objetivo é selecionar ou recomendar genótipos para o plantio, o estudo da interação entre genótipo x ambiente é de extrema importância. Entretanto, tal estudo não fornece informações pormenorizadas sobre o comportamento de cada cultivar diante das variações ambientais. Assim, tornam-se necessárias as análises de adaptabilidade e de estabilidade para a identificação e recomendação de materiais superiores em diferentes ambientes. Embora a literatura apresente diversos métodos, para realização da análise de adaptabilidade e estabilidade, nenhum leva em consideração a presença de fenótipos não normais, ou seja, distribuições de valores fenotípicos assimétricos ou com caudas pesadas. Desta forma, caso haja a presença desse tipo de valores fenotípicos, os métodos podem sofrer a influência de modo que a recomendação pode ser errônea, ou seja, o uso de tais métodos ocasionam estimativas inadequadas, que não refletem a verdadeira relação existente entre a variação ambiental e a resposta fenotípica. Uma solução interessante para tratar este problema de maneira unificada, isto é, a presença de pontos discrepantes ou assimetria, é a utilização de regressão quantílica (RQ). Tal metodologia, diferentemente dos métodos de regressão usuais, que utilizam a média condicional para explicar a relação funcional entre a variação ambiental e a resposta fenotípica, faz uso de funções quantílicas condicionais. Desta forma, a RQ possibilita escolher o quantil que melhor representa a relação funcional de interesse com o intuito de contemplar naturalmente a mencionada falta de normalidade. Desta forma, o presente trabalho tem como objetivo apresentar a metodologia de regressão quantílica, através de uma discussão detalhada de seus fundamentos teóricos, evidenciando, com aplicações concretas, seu uso em análise de adaptabilidade e estabilidade, fornecendo assim um material de fácil acesso para leitores interessados no assunto, contribuindo com pesquisadores e interessados nesta área. Para avaliação da técnica foram simulados valores fenotípicos, com distribuições simétrica, simétrica com *outliers*, assimétrica à direita, assimétrica à direita com *outliers*, assimétrica à esquerda e assimétrica à esquerda com *outliers*. Além disso, foram utilizados dados provenientes de um

experimento sobre produção de matéria seca de 92 genótipos de alfafa (*Medicago sativa*) avaliados em 20 ambientes. Sugere-se que, para valores fenotípicos simétricos deve-se averiguar se este possui *outlier*, se sim é utilizada ou a regressão não paramétrica ou a RQ ( $\tau = 0,50$ ), se não, se utiliza ou a metodologia de Eberhart e Russell (1966) ou a RQ ( $\tau = 0,50$ ). Já se o fenótipo for assimétrico, com ou sem a presença de *outlier*, utiliza-se RQ ( $\tau = 0,25$ ) para assimetria a direita e RQ ( $\tau = 0,75$ ) para assimetria à esquerda. De acordo com os resultados encontrados a RQ foi eficiente para classificação de genótipos de alfafa.

## ABSTRACT

BARROSO, Laís Mayara Azevedo, M.Sc., Universidade Federal de Viçosa, February, 2014. **Quantile Regression in the evaluation of adaptability and phenotypic stability.** Advisor: Moysés Nascimento. Co-Advisors: Ana Carolina Campana Nascimento, Cosme Damião Cruz and Leonardo Lopes Bhering.

In plant breeding, when the objective is to select or recommend genotypes to be planted, the study of the interaction between genotype and environment plays an important role. However, this kind of study does not provide detailed information on the behavior of each cultivar due to environmental variations. Thus, it becomes necessary to perform analyses of stability and adaptability for identification and recommendation of superior materials in different environments. Although the literature presents several methods for performing analysis of adaptability and stability, none of them take account of the presence of non-normal phenotype, in other words, phenotypic values asymmetric distributions or heavy tails. Thus, if there is the presence of such phenotypic values, the methods can be influenced and the recommendation may be mistaken, that is, the use of such methods cause inadequate estimates that do not reflect the true relationship between the variation environmental and phenotypic response. An interesting solution for treating this problem in a unified way, that is, the presence of outliers or asymmetry is to use the quantile regression (QR). Such methodology, besides the usual regression methods, using the conditional mean to explain the functional relationship between environmental variation and phenotypic response, makes use of conditional quantile functions. This way the QR is possible to choose the quantile which best represents the functional relationship of interest in order to naturally cover the lack of normality cited above. Thus, this paper aims to present the methodology of quantile regression, through a detailed discussion of its theoretical foundations, demonstrating it by concrete applications, its use in analysis of adaptability and stability, thus providing an easily and accessible material for readers interested in that subject, contributing researchers and those interested in this area. To the technique evaluation symmetric distributions phenotypic values, symmetric with outliers, right asymmetric, right asymmetric with outliers, left asymmetric and left asymmetric with outliers were simulated. Furthermore, we used data from an experiment on dry matter yield of 92 genotypes of alfalfa (*Medicago sativa*) evaluated in 20 environments. It is suggested that, for symmetrical phenotypic values should be determined if it has outlier, if it has a QR ( $\tau = 0,50$ )

should be used, if not, should be used either Eberhart and Russell methodology (1966) or QR ( $\tau = 0,50$ ). Since the phenotype is asymmetric, with or without the presence of outlier, it uses QR ( $\tau = 0,25$ ) to right asymmetry and QR ( $\tau = 0,75$ ) to the left asymmetry. According to the results the QR method was efficient for classifying alfalfa genotypes.

## 1. INTRODUÇÃO GERAL

No melhoramento genético vegetal, seja na fase de seleção e, principalmente, na fase de recomendação, o conhecimento do componente de interação genótipo x ambiente (IGA) é de extrema importância. Entretanto, tal conhecimento não fornece informações detalhadas a respeito do comportamento de cada cultivar frente às variações ambientais (CRUZ et al., 2012). Assim, visando identificar e recomendar materiais superiores em diferentes ambientes, o estudo pormenorizado desta componente, por meio de análises de adaptabilidade e de estabilidade, torna-se necessário.

As metodologias de análise de adaptabilidade e estabilidade são diferentes quanto aos conceitos adotados e a certos princípios estatísticos empregados (CRUZ et al., 2012). Como exemplo, pode-se citar as metodologias paramétricas baseadas em regressão (FINLAY e WILKINSON, 1963; EBERHART e RUSSELL, 1966; CRUZ et al., 1989; NASCIMENTO et al., 2010; NASCIMENTO et al., 2011) e análises não paramétricas, tais como os métodos propostos por Lin e Binns (1988), Carneiro (1998) além do método centroide (ROCHA et al. 2005) e seus desenvolvimentos posteriores, centroides múltiplos e ampliado (NASCIMENTO et al., 2009a, 2009b).

Dentre as metodologias citadas, uma classe que apresenta grande destaque é a formada pelos métodos baseados em regressão. A utilização dessas metodologias pode ser verificada no trabalho de Léo et al. (2005), os quais classificaram cultivares de alfafa quanto a adaptabilidade e estabilidade e, nos estudos de Pereira et al. (2009) e Nascimento et al. (2011), em que foram avaliadas cultivares de feijão e café, respectivamente.

Embora tais métodos sejam úteis, devido ao processo de estimação, o qual utiliza de médias condicionais para explicar a relação entre a variação ambiental e a resposta fenotípica, os mesmos sofrem a ação de pontos extremos e são pouco representativos quando a distribuição do fenótipo não é normal. Visando contornar este problema, Nascimento et al. (2010) apresentam uma solução para tratar fenótipos que não possuem distribuição normal. Especificamente, tal metodologia é recomendada quando a distribuição do fenótipo apresenta caudas pesadas, ou seja, distribuições que podem gerar *outliers*. Nesta metodologia, visto que para a obtenção dos estimadores utilizam-se medianas, as estimativas dos parâmetros são menos sensíveis a pontos

extremos que os estimadores clássicos obtidos pelo método dos mínimos quadrados ordinários, que são baseados em médias. Na prática, a presença de *outliers* pode proporcionar estimativas inadequadas, que não refletem a verdadeira relação existente entre a variação ambiental e a resposta fenotípica, e assim superestimar ou subestimar o parâmetro de adaptabilidade em métodos baseados em regressão.

Distribuições com caudas pesadas são na verdade um caso particular de distribuições de fenótipos que não apresentam normalidade. Assim, tanto para casos em que a distribuição do fenótipo apresente assimetria quanto caudas pesadas, a utilização de métodos usuais, baseados em regressão, os quais assumem uma relação funcional entre a variação ambiental e a resposta fenotípica através da média condicional ou mediana, podem proporcionar estimativas menos acuradas, que não refletem a verdadeira relação existente entre a variação ambiental e a resposta fenotípica. Outras metodologias tais como as propostas por Lin e Binns (1988) e Carneiro (1998), também podem ocasionar resultados impróprios na presença de fenótipos não normais, como por exemplo, na presença de pontos extremos, o valor dos parâmetros pode ser inflacionado e provocar a incorreta classificação do genótipo quanto à adaptabilidade e estabilidade. Quando se consideram as metodologias dos centroides múltiplos e centroide ampliado (NASCIMENTO et al., 2009a, 2009b), baseadas em componentes principais, a presença de pontos discrepantes influencia diretamente a configuração gráfica obtida por meio dos escores dos genótipos.

Uma solução interessante, não utilizada até o momento em estudos de adaptabilidade e estabilidade para tratar os problemas ocasionados pela presença de *outliers* ou assimetria na distribuição fenotípica é a utilização de regressão quantílica (RQ), que diferentemente dos métodos usuais, baseados em regressão, utiliza quantis condicionais e não a média para explicar a relação funcional entre a variação ambiental e a resposta fenotípica. Assim, a RQ possibilita escolher o quantil que melhor representa relação entre a variação ambiental e a resposta fenotípica com o intuito de contemplar naturalmente os problemas mencionados acima.

A RQ já foi abordada no melhoramento genético vegetal em alguns trabalhos, como por exemplo, De La Veja e Chapman (2010) e Gourджи et al. (2012). De La Veja e Chapman (2010) utilizaram a RQ para estimar o ganho genético como função limite superior da relação entre os rendimentos híbridos e ano de lançamento. Gourджи et al. (2012) utilizaram a RQ para ajustar os rendimentos observados para mudanças nos

locais e condições ambientais de ensaios ao longo do tempo, um passo considerado necessário para avaliar verdadeiros ganhos genéticos em condições teoricamente constantes.

Silva e Porto Júnior (2006) usaram a RQ para fornecer uma visão mais detalhada dos impactos gerados pelo sistema financeiro na distribuição condicional do crescimento econômico. Beyerlein et al. (2011) utilizou RQ em análises GWAS (*Genome-Wide Association Study*) na área de medicina humana e enfatizou vantagens estatísticas e biológicas ao se estimar efeitos de marcadores em diferentes quantis das distribuições dos fenótipos. Nascimento et al. (2012) utilizaram a RQ para investigar os determinantes da eficiência de produtores de leite de MG, para diferentes níveis de eficiência.

Diante do exposto, este trabalho pretende apresentar a metodologia da Regressão Quantílica e seus aspectos teóricos e, além disso, apresentar uma proposta de análise em estudos de adaptabilidade e estabilidade.

De acordo com o apresentado, este trabalho está organizado em revisão de literatura, capítulo 1 e conclusões.

Na revisão de literatura apresentam-se métodos de adaptabilidade e estabilidade baseado em regressão linear simples (EBERHART e RUSSELL, 1966) e baseado em análises não paramétricas (NASCIMENTO et al., 2010). Posteriormente, é tratada a Regressão Quantílica, seus conceitos, modelo, processo de estimação, construção de intervalos de confiança e coeficiente de determinação.

No capítulo 1 é apresentada uma aplicação da RQ, em que os autores propõem uma nova metodologia de análise de adaptabilidade e estabilidade fenotípica de genótipos de alfafa baseada em RQ.

Finalmente, apresentam-se as conclusões do trabalho.

## 2. REVISÃO DE LITERATURA

### 2.1. Interação Genótipo x Ambiente (IGA)

Para o estudo de um determinado ambiente, tem-se que a manifestação fenotípica é o efeito da ação do genótipo considerando a influência do meio (CRUZ et al., 2012). Todavia, quando é avaliado um conjunto de ambientes, além dos efeitos genotípicos e ambientais, deve-se considerar o efeito adicional ocasionado pela interação destes efeitos.

Esta interação é importante, pois, caso exista, pode ocorrer de genótipos superiores em um determinado ambiente não os serem em outro. Tal fato influencia o ganho de seleção e dificulta a recomendação de cultivares com ampla adaptabilidade (CRUZ et al., 2012). Desta forma a avaliação da magnitude e significância da IGA torna-se imprescindível em um programa de melhoramento.

A quantificação do componente da variação atribuído aos efeitos da IGA ( $\sigma_{ga}^2$ ) é realizada por meio da análise conjunta da variância, através de um sistema no qual se igualam os quadrados médios aos respectivos estimadores dos componentes de suas esperanças matemáticas. Posteriormente, por meio do teste F, é verificada a significância deste componente.

Dada à significância do componente de variância da IGA é necessário metodologias para contornar alguns inconvenientes, como o fato de um genótipo ser o melhor em um ambiente e não o ser em outro. Uma maneira de contornar estes inconvenientes proporcionados pela IGA é a utilização da estratificação da região considerada em sub-regiões com características ambientais semelhantes. Entretanto, ao utilizar este procedimento uma fração da interação ainda permanece, pois há ocorrência de fatores incontroláveis dos ambientes, como chuvas e temperaturas (CRUZ et al., 2012). Outra maneira é utilizar estudos de adaptabilidade e estabilidade, os quais permitem identificar cultivares que apresentam comportamento previsível e que respondam de forma satisfatória às variações dos ambientes.

Existem na literatura diversas metodologias que fazem estudos de adaptabilidade e estabilidade. É válido ressaltar que, para a escolha do método, devem-se levar em consideração alguns fatores, como o número de ambientes avaliados. Conforme descrito

em Resende et al. (2001) quando as metodologias são baseadas unicamente em análises de variância, estas são recomendadas quando se dispõe de um número restrito de ambientes, geralmente entre 3 e 5, enquanto que as metodologias que são baseadas em análises de variância e de regressão necessitam, para a sua aplicação, de um número não muito grande de ambientes de 5 ou mais.

Dentre as metodologias existentes podem-se citar o método Tradicional (apud CRUZ et al., 2012, p. 143), o método proposto por Plaisted e Peterson (1959), a metodologia de Finlay e Wilkinson (1963), o método de Eberhart e Russell (1966), a metodologia proposta por Tai (1971), o método de Cruz, Torres e Vencovsky (1989). Além destes, pode-se citar as metodologias de Lin e Binns (1988), regressão não paramétrica (NASCIMENTO et al., 2010), método centroide (ROCHA et al., 2005) e seus desenvolvimentos posteriores, centroides múltiplos e ampliado (NASCIMENTO et al., 2009a, 2009b), além das metodologias baseadas em análises gráficas, tais como AMMI (GAUCH JUNIOR, 2006) e GGE biplot (YAN e TINKER, 2006). Estas metodologias distinguem-se dos conceitos de adaptabilidade e estabilidade adotados e dos princípios estatísticos empregados (CRUZ et al., 2012).

A seguir são apresentadas algumas metodologias de análise de adaptabilidade e estabilidade utilizadas neste trabalho.

## 2.2. Metodologia de Adaptabilidade e Estabilidade baseada em Regressão Linear simples

### 2.2.1. Eberhart e Russell (1966)

Nos métodos de adaptabilidade e estabilidade baseados em análise de regressão, a variação de cada genótipo frente às variações ambientais é estabelecida por meio do ajuste de um modelo de regressão linear simples.

Dessa forma, para um experimento com  $g$  genótipos e  $a$  ambientes define-se o seguinte modelo estatístico:

$$Y_{ij} = \beta_{0i} + \beta_{1i}I_j + \psi_{ij}, \quad (1)$$

em que:  $Y_{ij}$  é a média de genótipo  $i$ , obtida por meio de  $r$  repetições, no ambiente  $j$ , com  $i=1,\dots,g$  e  $j=1,\dots,a$ ;  $\beta_{0i}$  é a constante da regressão referente ao  $i$ -ésimo genótipo;  $\beta_{1i}$  é o coeficiente de regressão, que mede a resposta do  $i$ -ésimo genótipo à variação do ambiente;  $I_j$  é o índice ambiental codificado  $I_j = \frac{\sum_i Y_{ij}}{g} - \frac{\sum_i \sum_j Y_{ij}}{ga}$ ;  $\psi_{ij} = \delta_{ij} + \bar{\varepsilon}_{ij}$  é o erro aleatório composto pela soma do desvio da regressão  $\delta_{ij}$  e o erro experimental médio  $\bar{\varepsilon}_{ij}$ .

As estimativas do índice ambiental  $I_j$ , indicam a qualidade do ambiente, uma vez que valores negativos de  $I_j$  identificam ambientes desfavoráveis e valores positivos de  $I_j$ , ambientes favoráveis.

Na metodologia de Eberhart e Russell (1966), o conceito de adaptabilidade refere-se à capacidade de os genótipos responderem ao estímulo do ambiente enquanto que o conceito de estabilidade refere-se à capacidade de os genótipos mostrarem um comportamento previsível em função do estímulo do ambiente (CRUZ et al., 2012).

Para obtenção dos parâmetros de um modelo de regressão, utiliza-se o método de mínimos quadrados ordinários (MQO). Este método consiste em tomar como estimativas dos parâmetros, os valores que minimizem a soma de quadrados dos desvios (erros). Assim, os estimadores de mínimos quadrados dos coeficientes do modelo (1)

são dados por:  $\hat{\beta}_{0i} = \bar{Y}_i$  e  $\hat{\beta}_{1i} = \frac{\sum_j Y_{ij} I_j}{\sum_j I_j^2}$ .

Nesta metodologia, além de  $\beta_{1i}$ , o desvio da regressão ( $\sigma_{d_i}^2$ ) também é considerado como parâmetro para a classificação dos genótipos. Assim, os parâmetros de adaptabilidade e estabilidade são dados pelo coeficiente angular da regressão  $\beta_{1i}$  e o desvio da regressão  $\sigma_{d_i}^2$ , respectivamente. Os autores consideram como ideal o genótipo que apresenta  $\hat{\beta}_{1i} = 1$  e desvios da regressão tão pequenos quanto possíveis.

A classificação dos genótipos quanto à adaptabilidade é feita a partir das estimativas deste parâmetro. Valores de  $\hat{\beta}_{1i}$  maiores que 1 indicam que os genótipos

são recomendados a ambientes favoráveis, enquanto que, para valores de  $\hat{\beta}_{1i}$  menores que 1 os genótipos são indicados a ambientes desfavoráveis. Já para o caso onde  $\hat{\beta}_{1i}$  é igual a 1, os genótipos são classificados como de adaptabilidade geral. A avaliação da hipótese de que o parâmetro seja igual à unidade,  $H_{0i} : \beta_{1i} = 1$ , é realizada por meio do teste t, cuja estatística é dada por  $t = \frac{\hat{\beta}_{1i} - 1}{\sqrt{\hat{V}(\hat{\beta}_{1i})}}$ , em que  $\hat{V}(\hat{\beta}_{1i}) = \frac{QMR}{r \sum_j I_j^2}$ . Esta estatística está associada ao número de graus de liberdade do resíduo da análise de variância conjunta e ao nível de significância  $\alpha$ .

O parâmetro de estabilidade ( $\sigma_{d_i}^2$ ) é obtido por meio de,  $\hat{\sigma}_{d_i}^2 = \frac{QMD_i - QMR}{r}$ , em que  $QMD_i$ : quadrado médio do desvio do genótipo i;  $QMR$ : quadrado médio do resíduo; r: número de repetições.

A avaliação da hipótese  $H_0 : \sigma_{d_i}^2 = 0$  é realizada por meio do teste F, cuja estatística é dada por  $F = \frac{QMD_i}{QMR}$ , que está associada a  $a - 2$  e  $m$  graus de liberdade, em que  $a$  é o número de ambientes,  $m$  é o número de graus de liberdade do resíduo da análise conjunta e a um nível de significância  $\alpha$ .

Em relação ao parâmetro de estabilidade ( $\sigma_{d_i}^2$ ), genótipos com  $\sigma_{d_i}^2$  igual a zero classificam-se como de alta previsibilidade, ou seja, os genótipos mostram um comportamento altamente previsível em função do estímulo do ambiente, e genótipos com  $\sigma_{d_i}^2$  maior que zero classificam-se como de baixa previsibilidade.

Além disso, o valor de  $\hat{\beta}_{0i} = \bar{Y}_i$  representa a média de produção do genótipo i quando se considera todos os ambientes, assim, quanto maior o valor de sua estimativa maior o interesse para o plantio.

### 2.3. Metodologia de Adaptabilidade e Estabilidade baseada em análise não paramétrica

### 2.3.1. Adaptabilidade e estabilidade via regressão não paramétrica (NASCIMENTO et al., 2010)

Nesta metodologia, da mesma forma que em Eberhart e Russell (1966), a adaptabilidade é definida como a capacidade dos genótipos responderem ao estímulo do ambiente, sendo que os mesmos são classificados de acordo com o resultado do teste t aplicado ao coeficiente de regressão.

A definição de estabilidade também é a mesma que a descrita no método de Eberhart e Russell (1966). Entretanto, nesta metodologia, diferentemente de Eberhart e Russell (1966), utiliza-se o coeficiente de determinação,  $R_{T_i}^2$  como parâmetro de estabilidade, de modo que os genótipos são classificados da seguinte forma: i) genótipos com estabilidade ou previsibilidade alta, são aqueles em que  $R_{T_i}^2$  são maiores que 70% e, ii) genótipos com estabilidade ou previsibilidade baixa, são aqueles em que  $R_{T_i}^2$  são menores que 70%.

Segundo Nascimento et al. (2010) a estimação do coeficiente angular  $\beta_{li}$ , parâmetro de adaptabilidade via regressão não paramétrica, é realizada da seguinte forma: calculam-se todos os valores amostrais  $A = [a(a-1)]/2$  de  $S_{ikl} = (Y_{il} - Y_{ik}) / (I_l - I_k)$  para todo  $i = 1, \dots, g$  e  $1 \leq k < l \leq a$ , em que  $Y_{il}$  é a média referente ao i-ésimo genótipo no l-ésimo ambiente;  $I_l$  é o valor do índice ambiental referente ao l-ésimo ambiente. Assim, o estimador de  $\beta_{li}$  é dado por  $\hat{\beta}_{li} = \text{mediana} \{S_{ikl}, 1 \leq k < l \leq a\}$ . Então, se  $A$  é ímpar,  $A = 2k + 1$ , tem-se  $\hat{\beta}_{li} = S^{k+l}$ . Se  $A$  é par,  $A = 2k$ , tem-se  $\hat{\beta}_{li} = (S^k + S^{k+l}) / 2$ .

A avaliação da hipótese  $H_0 : \beta_{li} = 1$  é realizada da mesma maneira que na metodologia de Eberhart e Russell (1966), através do teste t.

O intercepto  $\beta_{0i}$  é estimado, segundo Theil (1950), a partir da mediana de todos  $\hat{\beta}_{0i} = y_{ij} - \hat{\beta}_{li} I_j$ .

Nesta metodologia a estimação da estabilidade é feita a partir do cálculo do coeficiente de determinação, isto é, a variabilidade da variável dependente que é explicada pelo modelo ajustado e é dada por:

$$R_{T_i}^2 = \text{Var}(\hat{y}_{ij}) / \text{Var}(y_{ij}) \quad (2)$$

em que  $\text{Var}(\hat{y}_{ij})$  é a variância dos valores estimados e  $\text{Var}(y_{ij})$  é a variância dos valores observados.

Para utilizar o método proposto por Nascimento et al. (2010), é necessário saber se há um ponto extremo que esteja subestimando ou superestimando o parâmetro de adaptabilidade. Desta forma, definiu-se como medida da influência de um ponto a variação, em módulo, entre os estimadores do coeficiente de inclinação estimados pelos métodos de mínimos quadrados ( $\hat{\beta}_{li}^*$ ) e pelo método de regressão não paramétrica ( $\hat{\beta}_{li}$ ) para cada genótipo, isto é,  $\Delta\hat{\beta}_{li} = |\hat{\beta}_{li}^* - \hat{\beta}_{li}|$ . Os autores consideram o valor de 0,05 de variação para assumir a existência de um ponto extremo. Caso a variação não ultrapasse este valor definido, é utilizado o método de Eberhart e Russell (1966).

Esta metodologia é menos influenciada por pontos extremos e caso os genótipos tenham resposta diferenciada em algum ambiente, a má interpretação do parâmetro de adaptabilidade é evitada.

Diversos autores eliminam os *outliers* para contornar os inconvenientes provocados por eles (LY et al., 2013), entretanto ao realizar esta ação pode acontecer de haver perda de informação no conjunto de dados, pois mesmo interferindo na estimação, o *outlier* faz parte do fenômeno estudado.

## 2.4. Regressão Quantílica

### 2.4.1. Introdução

Uma das técnicas mais estudadas no meio acadêmico é a Regressão Linear (RL), cujo principal objetivo é estabelecer uma relação funcional entre a variável resposta e a variável preditora. No modelo de regressão linear simples (RLS), a relação entre a

variável preditora  $X$  e a variável resposta  $Y$  é dada pela equação de uma reta e é representada por:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad (3)$$

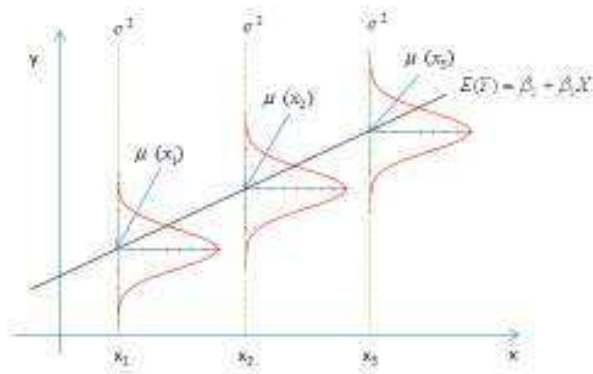
em que o intercepto  $\beta_0$  e o coeficiente angular  $\beta_1$  são constantes desconhecidas, também conhecidos como parâmetros e  $e_i$  são os erros aleatórios. Os termos de erros têm média zero, variância desconhecida  $\sigma^2$  e são homocedásticos. Adicionalmente assumimos que os erros são não correlacionados e normalmente distribuídos.

Uma maneira de se estimar os coeficientes do modelo (3) é através do método dos mínimos quadrados ordinários (MQO), no qual se adotam como estimativas dos parâmetros os valores que minimizam a soma de quadrados dos erros, ou seja, das diferenças entre as distâncias verticais das observações  $y_i$  e a reta estimada (GUJARATI, 2011; CECON et al., 2012).

Apesar da facilidade de interpretação e implementação, a RLS apresenta algumas limitações. Em primeiro lugar, os pressupostos do modelo nem sempre são válidos para conjuntos de dados reais. A pressuposição de homocedasticidade (Figura 1) geralmente falha. Segundo Gujarati (2011) existem várias razões que podem causar a quebra do pressuposto de homocedasticidade dos erros e, dentre elas podem-se citar: modelos de aprendizagem com o erro, aperfeiçoamento de técnicas de coleta de dados, dados discrepantes, entre outras. Desta forma, quando excluimos o pressuposto de homocedasticidade, embora os estimadores de mínimos quadrados sejam ainda não tendenciosos os mesmos não possuem variância mínima (MONTGOMERY et al., 2012). Assim, os estimadores obtidos por meio do método dos mínimos quadrados não serão BLUE (*Best Linear Unbiased Estimator*).

Quando a distribuição apresenta caudas pesadas, a média condicional pode se tornar uma medida inadequada e enganosa de localização central, pois é fortemente influenciada por *outliers*.

Além disso, o cálculo dos p-valores se baseiam no pressuposto de normalidade e caso exista violação desta condição pode ocorrer viés nos p-valores, acarretando em testes de hipóteses inválidos (HAO E NAIMAN, 2007).



**Figura 1.** Representação do modelo RLS com erros homocedásticos.

Para contornar a limitação referente à distribuição normal dos erros, pode-se utilizar o método de minimização dos erros absolutos, uma vez que este é robusto na presença de *outliers* e descreve melhor uma medida de posição central da distribuição condicional da variável resposta, uma vez que estima o valor mediano da distribuição.

Koenker e Bassett (1978) propuseram uma abordagem mais geral, a regressão quantílica (RQ), que se baseia no método dos erros absolutos ponderados. Entretanto, nesta metodologia não se considera apenas o valor mediano, e sim é realizada uma ponderação na minimização dos erros para se estimar os diversos quantis de interesse.

A utilização de diversos quantis possibilita a obtenção de maiores informações de localização do que quando se utiliza apenas o centro da distribuição. Desta forma é possível examinar uma localização na cauda inferior (por exemplo, o quantil 0,1) ou na cauda superior (por exemplo, o quantil 0,9) quando o pesquisador necessita de informações sobre subpopulações específicas (HAO e NAIMAN, 2007). Como por exemplo, nos estudos envolvendo problemas econômicos (SILVA e PORTO JUNIOR, 2006), medicina humana (BEYERLEIN et al., 2011) e problemas sociais (HAO e NAIMAN, 2007).

Além disso, de acordo com Koenker (2005), modelos de RQ são capazes de incorporar uma possível heterocedasticidade, que seria detectada a partir da variação das estimativas dos coeficientes dos parâmetros para diferentes quantis ( $\tau$ 's).

#### 2.4.2. Quantil como solução para um problema de minimização

De acordo com Hao e Naiman (2007), um quantil de ordem  $\tau$  pode ser considerado como a solução de um determinado problema de minimização.

O quantil de ordem  $\tau$  de uma população ou de uma amostra é o valor  $m$  tal que  $100\tau$  dos valores populacionais ou amostrais são inferiores a ele, com  $0 < \tau < 1$ .

Outra definição bastante utilizada está descrita em Hao e Naiman (2007), onde os autores falam que o  $\tau$ -ésimo quantil da função de distribuição acumulada  $F$  é o valor mínimo de  $y$  tal que  $F(y) \geq \tau$ .

Seja  $Y$  uma variável aleatória com função de distribuição acumulada  $F$ . Hao e Naiman (2007) mostraram que o valor de  $m$  que minimiza a média da distancia absoluta,  $E|Y - m|$ , é a mediana, ou seja, o ponto em que a derivada em relação à  $m$  é nula ou onde as duas derivadas direcionais mudam de sinal. A demonstração deste resultado encontra-se no Apêndice C.

Considerando uma amostra também é possível definir a distância média absoluta de  $m$  até pontos amostrais a partir de  $f(m) = \frac{1}{n} \sum_{i=1}^n |y_i - m|$ . Desta forma, a função será mínima quando a derivada valer  $-\frac{1}{n}$ , para  $m < y_i$  e  $\frac{1}{n}$  para  $m > y_i$ . Além disso, a função não é diferenciável em  $m = y_i$ , e então, assumirá uma derivada direcional de  $-\frac{1}{n}$  na direção negativa e  $\frac{1}{n}$  no sentido positivo.

Este resultado pode ser generalizado para qualquer quantil de interesse,  $\tau \in [0, 1]$ . Para tanto, considere a distância absoluta de  $Y$  para um dado  $p$  aplicado diferentes pesos se  $Y$  está à esquerda ou à direita de  $p$ , ou seja:

$$d_{\tau}(Y, p) = \begin{cases} (1 - \tau)|Y - p|, & \text{se } Y < p \\ \tau|Y - p|, & \text{se } Y \geq p \end{cases} \quad (4)$$

Como apresentado em Hao e Naiman (2007), o mínimo da média desta função ( $E[d_{\tau}(Y, p)]$ ) ocorre quando  $p$  é o  $\tau$ -ésimo quantil. Da mesma forma, uma amostra do  $\tau$ -ésimo quantil é o valor de  $p$  que minimiza a distância ponderada média:

$$\frac{1}{n} \sum_{i=1}^n d_{\tau}(y_i, p) = \frac{1 - \tau}{n} \sum_{y_i < p} |y_i - p| + \frac{\tau}{n} \sum_{y_i \geq p} |y_i - p| \quad (5)$$

em que  $d_\tau(y_i, p)$  é a distância de  $y_i$  até  $p$ ,  $n$  é o tamanho amostral,  $\tau$  é o quantil de interesse ( $\tau \in [0,1]$ ),  $p$  é um ponto qualquer e  $y_i$  são os valores observados.

### 2.4.3. Modelo da RQ

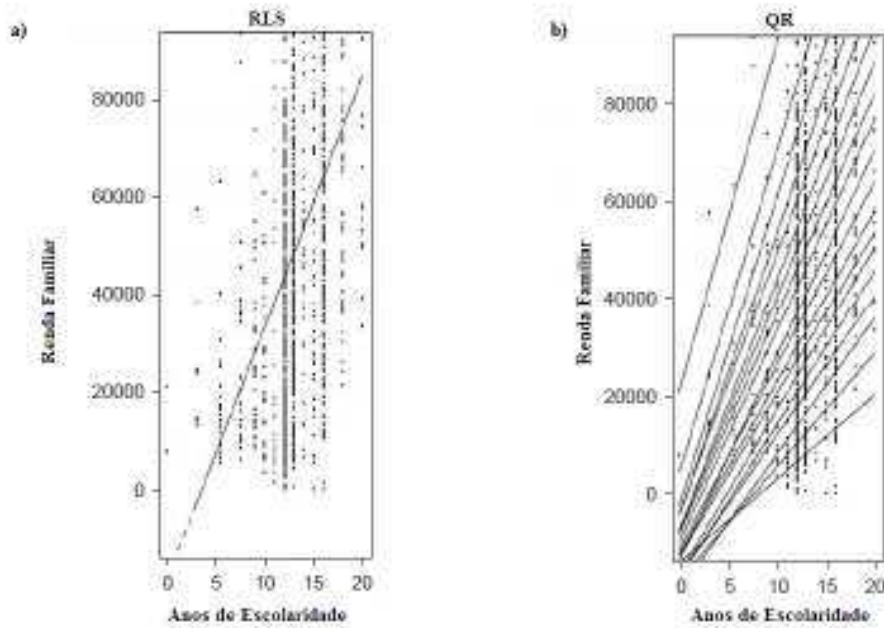
O modelo da Regressão Quantílica, que descreve a relação funcional entre a variável dependente e a variável independente, pode ser descrito como:

$$Y_i = \beta_{0i}(\tau) + \beta_{1i}(\tau)X_i + e_i(\tau) \quad (6)$$

em que  $\beta_{0i}(\tau)$  é a constante da regressão;  $\beta_{1i}(\tau)$  é o coeficiente da regressão;  $e_i(\tau)$  são os erros aleatórios independentes e identicamente distribuídos com quantil de ordem  $\tau$  igual a zero;  $X$  é variável independente e  $\tau$  refere-se ao quantil assumido ( $\tau \in [0,1]$ ).

Nos modelos de RLS (3) estima-se apenas uma reta para explicar todo o conjunto de dados. Entretanto, a linha da regressão não captura mudanças na distribuição da variável dependente. Já na RQ (6) podem ser estimadas retas para cada quantil de interesse, desta maneira se torna mais adequado à interpretação dos resultados para o conjunto de dados com presença de assimetria, pois através dela é possível traçar a relação em regiões centrais, através da mediana, e nas caudas da distribuição condicional de acordo com o interesse.

Pode-se observar que a RLS fornece apenas informações em termos médios da variável dependente (Figura 2A). Enquanto que a RQ possibilita observar a relação funcional em diferentes níveis da variável dependente (Figura 2B).



**Figura 2.** Ajuste de um modelo linear e diversos ajustes da regressão quantílica.  
 Fonte: Hao e Naiman (2007).

#### 2.4.4. Estimação

Conforme discutido anteriormente, tem-se que a estimação da RLS é feita por meio do MQO, no qual se deseja minimizar a soma de quadrados dos erros.

A principal diferença entre as estimações da RQ e da RLS é que a distância dos pontos observados a reta estimada na RQ é medida minimizando a média ponderada da soma das distâncias verticais, sendo que para pontos abaixo da linha atribui-se peso  $1 - \tau$ , e para pontos acima da linha peso  $\tau$ , conforme apresentado em Hao e Naiman (2007).

Cada escolha do valor do peso  $\tau$  resultará numa função ajustada do quantil condicional. Assim, como descrito na seção 2.4.2, tem-se que um quantil  $\tau$  pode ser visto como o ponto em que minimiza a distância média ponderada, com pesos dependendo da localização do ponto, se o ponto está acima ou abaixo da reta ajustada.

O objetivo então é encontrar os valores de  $\beta'_i$ s que minimizem a equação:

$$\sum_{i=1}^n d_{\tau}(y_i, \hat{y}_i) = \tau \sum_{y_i \geq \beta_{0i}(\tau) + \beta_{1i}(\tau)X_i} |y_i - \beta_{0i}(\tau) - \beta_{1i}(\tau)X_i| + (1 - \tau) \sum_{y_i < \beta_{0i}(\tau) + \beta_{1i}(\tau)X_i} |y_i - \beta_{0i}(\tau) - \beta_{1i}(\tau)X_i| \quad (7)$$

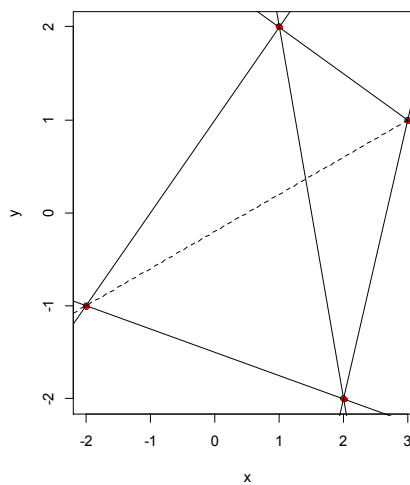
em que  $d_\tau$  é a distância entre  $y_i$  e  $\hat{y}_i$ ;  $\beta_{0i}(\tau)$  é a constante da regressão;  $\beta_{1i}(\tau)$  é o coeficiente da regressão;  $X$  é variável independente e  $\tau$  refere-se ao quantil assumido ( $\tau \in [0,1]$ ).

Ao minimizar a Equação 7 tem-se como resultados a reta da regressão do quantil de interesse.

Hao e Naiman (2007) ilustram o algoritmo de minimização da soma dos erros ponderados para obtenção dos estimadores da RQ, considerando quatro pares de pontos hipotéticos ((-2,-1); (1,2); (2,-2); (3,1)) e as seis retas que ligam cada um destes pares de pontos (Figura 3). Para este exemplo considere ainda o valor de  $\tau=0,5$ , ou seja, deseje-se encontrar a reta da regressão mediana que passa por um par de pontos do conjunto de modo que, metade dos pontos se encontram abaixo desta reta e a outra metade acima. Desta forma, metade dos resíduos é negativa e a outra metade positiva (HAO e NAIMAN, 2007). Assim, é necessário minimizar a equação:

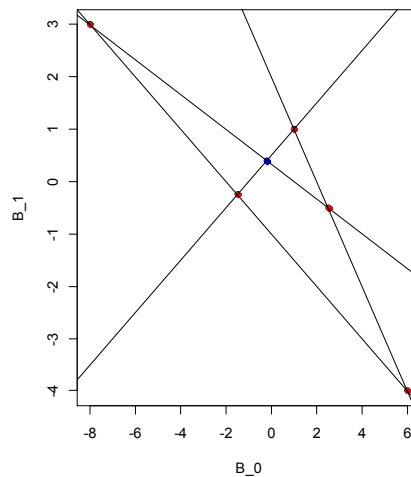
$$\sum_{i=1}^n d_\tau(y_i, \hat{y}_i) = 0,5 \sum_{y_i \geq \beta_{0i}(0,5) + X_i \beta_{1i}(0,5)} |y_i - \beta_{0i}(0,5) - X_i \beta_{1i}(0,5)| + (1-0,5) \sum_{y_i < \beta_{0i}(0,5) + X_i \beta_{1i}(0,5)} |y_i - \beta_{0i}(0,5) - X_i \beta_{1i}(0,5)|$$

A linha pontilhada é exatamente a linha ajustada pela regressão mediana (Figura 3), visto que a mesma apresenta exatamente um ponto abaixo e acima da reta e apresenta a menor soma das distâncias ponderadas considerando  $\tau=0,5$ .



**Figura 3.** Gráfico do plano (x,y)

Cada reta apresentada na Figura 3 pode ser representada por uma equação do tipo  $y = \beta_0 + \beta_1 x$ . Desta forma, podem-se obter seis pontos do tipo  $(\beta_0, \beta_1)$ . De posse destes pares, é possível estabelecer uma relação entre os pontos do plano  $(x, y)$  e as retas do plano  $(\beta_0, \beta_1)$ , uma vez que para o primeiro plano temos a equação da reta  $y = \beta_0 + \beta_1 x$ , e desta forma os pontos  $(\beta_0, \beta_1)$  encontram-se na reta  $\beta_1 = \left( \frac{y_i}{x_i} \right) - \left( \frac{1}{x_i} \right) \beta_0$ . Tal relação é denotada como dualidade ponto/reta (*point/line duality*) (EDGEWORTH, 1888) (Figura 4) (HAO E NAIMAN, 2007).



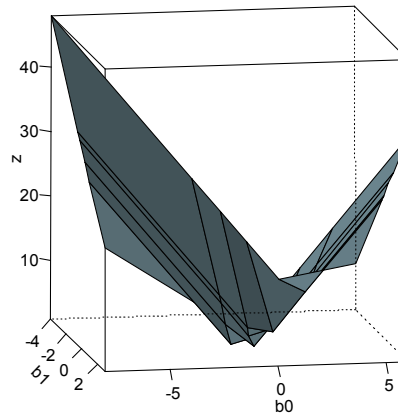
**Figura 4.** Gráfico do plano  $(\beta_0, \beta_1)$

A Figura 4 apresenta o plano  $(\beta_0, \beta_1)$  onde cada ponto representado corresponde a uma reta do plano  $(x, y)$ . O ponto em destaque (azul) corresponde à linha da regressão mediana ajustada na Figura 3.

Na Figura 4 as quatro linhas apresentadas correspondem aos quatro pontos da Figura 3. Estas linhas dividem o gráfico em regiões poligonais<sup>1</sup>. Os pontos que formam estas regiões correspondem a um conjunto de linhas no plano  $(x, y)$ . Além disso, estes pontos dividem o conjunto de dados da mesma forma. Portanto, como neste exemplo foram utilizadas apenas retas, tem-se que a função de  $(\beta_0, \beta_1)$  que minimiza a Equação 8 é linear em cada região. Considerando uma nova dimensão definida como os valores para todos os pares de betas e de

<sup>1</sup> Região Poligonal é a reunião de um polígono com o seu interior.

observações da equação 7 com  $\tau=0,5$  (Figura 5), conclui-se que esta função é convexa com um gráfico que forma uma superfície poliédrica<sup>2</sup> e o par de betas que minimiza a função é a reta mediana, ilustrada na Figura 3.



**Figura 5.** Superfície poliédrica.

Um algoritmo que permite minimizar a equação 7, estimando os coeficientes da regressão mediana  $(\hat{\beta}_0, \hat{\beta}_1)$ , é baseado em algoritmos de problemas de programação linear (HAO e NAIMAN, 2007). Tais algoritmos não fazem parte do escopo deste trabalho e assim será apresentada apenas a maneira como tais métodos trabalham. O método utilizado para solução de problemas de programação linear é o Método Simplex. A ideia da minimização no Método Simplex é que a partir de qualquer um dos pontos  $(\beta_0, \beta_1)$ , que corresponde aos vértices da região poligonal (Figura 5), é feita uma iteração percorrendo vértice a vértice das arestas da superfície e desta forma escolhe-se o caminho onde a distância é mínima. O principal objetivo neste método é obter o ponto que corresponde ao menor valor da equação 7.

Este resultado pode ser generalizado para qualquer estimador da regressão quantílica (KOENKER e D'OREY, 1987). Deste modo é necessário encontrar os estimadores  $\hat{\beta}_{0i}(\tau)$  e  $\hat{\beta}_{1i}(\tau)$  que minimizem a soma dos erros absolutos ponderados da equação 7, ou seja, deve-se minimizar a soma dos resíduos  $y_i - \hat{y}_i$  de maneira que resíduos com valores positivos recebem peso  $\tau$  e resíduos negativos recebem peso

<sup>2</sup> Superfície Poliédrica limitada convexa é a reunião de um número finito de polígonos planos e convexos.

$1 - \tau$ . Assim devem-se obter os estimadores  $\hat{\beta}_{0i}(\tau)$  e  $\hat{\beta}_{1i}(\tau)$  que minimizem a equação 7.

#### 2.4.5. Intervalo de Confiança

Na literatura são descritos dois métodos para construção do intervalo de confiança para os parâmetros dos modelos de RQ. O primeiro baseado em resultados assintóticos e o segundo baseado em *bootstrap*. Hao e Naiman (2007) consideram a utilização do procedimento baseado em *bootstrap* preferível, visto que o mesmo não depende de pressuposições quanto à distribuição dos dados.

Assim, neste trabalho será descrito apenas o método baseado em *bootstrap* para obter os intervalos de confiança desejados.

##### 2.4.5.1. Metodologia baseada em *bootstrap*.

De acordo com Koenker (2005) selecionam-se os pares de observações  $(y_i, x_i)$  com probabilidade  $1/n$ , onde  $n$  é o tamanho amostral. Desta forma constrói-se um novo vetor  $Y^*$  com os valores da variável resposta e uma nova matriz de delineamento  $X^*$ . Este procedimento é repetido  $M$  vezes e em cada repetição é calculado o vetor  $\hat{\beta}^*(\tau)$ . Com as  $M$  estimativas para o vetor de parâmetros do modelo da RQ é estimado o erro padrão de  $\hat{\beta}(\tau)$  a partir do erro padrão encontrado nas reamostragens.

Dessa forma um intervalo de confiança para  $\hat{\beta}(\tau)$ , com coeficiente de confiança  $\gamma = 1 - \alpha$ , é dado por:

$$\hat{\beta}_i(\tau) \pm z_{\alpha/2} s_{boot}, \quad (8)$$

em que  $z_{\alpha/2}$  é o quantil de ordem  $1 - \alpha/2$  da distribuição normal padrão e  $s_{boot}$  é o estimador do erro padrão do parâmetro  $\beta_i(\tau)$  encontrado por meio de procedimento *bootstrap*.

#### 2.4.6. Qualidade de ajuste na RQ

O coeficiente de determinação,  $R^2$ , mede a qualidade do ajuste de um modelo de regressão, ou seja, ele indica a proporção da variação da variável resposta que é explicada pela regressão.

Uma alternativa para calcular o coeficiente de determinação é dado por:

$$R^2(\tau) = \frac{Var(\hat{y}_{ij})}{Var(y_{ij})} \quad (9)$$

em que  $Var(\hat{y}_{ij})$  é a variância dos valores estimados e  $Var(y_{ij})$  é a variância dos valores observados.

Além disso, os valores de  $R^2(\tau)$  estão compreendidos entre 0 e 1 ( $0 \leq R^2(\tau) \leq 1$ ).

#### 2.4.7. Interpretação das estimativas dos parâmetros

##### 2.4.7.1. RLS

Considerando o exemplo utilizado por Hao e Naiman (2007) para o modelo de RLS, tem-se o coeficiente da regressão estimado ( $\hat{\beta}_1$ ) é \$6,314 para ED (anos de escolaridade do chefe da família), ou seja, o aumento de um ano na escolaridade aumenta a renda, na média, \$6,314.

Como na RLS tem-se como pressuposto a linearidade, observa-se que a mesma quantidade de aumento da média condicional poderia ocorrer por famílias em qualquer nível fixo de escolaridade. Por exemplo, para as famílias cujo chefe tem 9 ou 16 anos de escolaridade, mais de um ano de educação está associada a uma mesma quantidade de aumento na renda média (HAO e NAIMAN, 2007).

##### 2.4.7.2. RQ

Hao e Naiman (2007) afirmam que o modelo de RQ mais simples de se entender é a regressão mediana ( $\tau = 0,50$ ). Nesse caso a interpretação do coeficiente é realizada de maneira análoga ao de um coeficiente da RLS. Para uma covariável contínua, como

por exemplo, anos de escolaridade, a estimativa do coeficiente é interpretada como a alteração mediana da variável dependente correspondente para uma unidade de mudança da variável independente, ou seja, aumenta-se  $\hat{\beta}_1$  para os indivíduos que se encontram no quantil 0,50.

Da mesma forma que na RLS, o efeito sobre a resposta mediana de um aumento de um ano na educação é o mesmo para todos os níveis de ensino.

Considerando a Tabela 1, para a regressão mediana ( $\tau = 0,50$ ), tem-se que o valor do coeficiente para ED (\$4,794) é menor que o coeficiente da RLS (\$6,314), o que sugere que o aumento não seria tão substancial para a maior parte da amostra.

Hao e Naiman (2007) ajustaram 19 quantis para verificar os efeitos da educação em vários quantis da renda (Tabela 1). Observa-se que o aumento de um ano na educação provoca um aumento de \$1,130 na renda para o quantil 0,05 e \$1,782 para o quantil 0,10. Em outras palavras, os indivíduos que pertencem ao quantil 0,05 tem um aumento de \$1,130 na renda e os que pertencem ao quantil 0,10 apresentam um aumento de \$1,782 na renda.

Caso o interesse do pesquisador seja na cauda a direita tem-se que o aumento na renda do quantil 0,95 é muito maior que no quantil 0,90, passa de \$8,279 para \$9,575. Isto sugere a contribuição da alta escolaridade para a disparidade de renda (HAO e NAIMAN, 2007).

**Tabela 1.** Estimativas da RQ e seus erros padrão assintóticos para renda.

Quantil	ED	Quantil	ED
0,05	1,130 (36)	0,55	5,182 (86)
0,10	1,782 (41)	0,60	5,571 (102)
0,15	2,315 (51)	0,65	5,841 (107)
0,20	2,757 (51)	0,70	6,224 (129)
0,25	3,172 (60)	0,75	6,598 (154)
0,30	3,571 (61)	0,80	6,954 (150)
0,35	3,900 (66)	0,85	7,505 (209)
0,40	4,266 (73)	0,90	8,279 (316)
0,45	4,549 (82)	0,95	9,575 (605)
0,50	4,794 (92)		

Nota: erros padrão assintóticos estão em parênteses  
 Fonte: Hao e Naiman (2007).

## REFERÊNCIAS BIBLIOGRÁFICAS

BEYERLEIN, A.; VON KRIES, R., NESS, A. R., ONG, K. K. Genetic Markers of Obesity Risk: Stronger Associations with Body Composition in Overweight Compared to Normal-Weight Children. **PLoS ONE**. v. 6, 2011.

CARNEIRO, P.C.S. **Novas metodologias de análise da adaptabilidade e estabilidade de comportamento**. 168p. Tese (Doutorado) - Universidade Federal de Viçosa, Viçosa. 1998.

CECON, P. R.; SILVA, A. R.; NASCIMENTO, M.; FERREIRA, A. **Métodos estatísticos**. 1. ed., Viçosa: UFV, 229p., 2012.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 4.ed. Viçosa: UFV, v.1, 514p., 2012.

CRUZ, C. D.; TORRES, R. A.; VENCOVSKY, R. An alternative approach to the stability analysis proposed by Silva and Barreto. **Revista Brasileira de Genética**, v.12, p. 567-80., 1989.

DE LA VEJA, A. J.; CHAPMAN, S. C. Mega-Environment Differences Affecting Genetic Progress for Yield and Relative Value of Component Traits. **Crop Science**, v. 50, p. 574 - 583,2010.

EBERHART, S. A.; RUSSELL, W. A. Stability parameters for comparing varieties. **Crop Science**, v.6, p.36-40, 1966.

EDGEWORTH, F. On a new method of reducing observations relating to several quantiles. **Philosophical Magazine**, v. 25, p. 184-191, 1888.

FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, v.14, p.742-754, 1963.

GAUCH JUNIOR, H. G. Statistical analysis of yield trials by AMMI and GGE. **Crop Science**, Madison, v. 46, p. 488-1500, 2006.

GOURDJI, S. M.; MATHEWS, K. L. .; REYNOLDS, M.; CROSSA, J.; LOBELL, D.B. An assessment of wheat yield sensitivity and breeding gains in hot environments, *Proceedings of the royal society B*. v. 280, n. 1552, p. 1 – 8, 2012.

GUJARATI, D. N. **Econometria básica**. 5. ed. São Paulo: Pearson Makron Books, 920p., 2011.

HAO, L.; NAIMAN, D. Q. **Quantile regression**. Sage publications. 126p. 2007.

KOENKER, R. **Quantile Regression**. 1.ed. Cambridge University Press, v.1. 349p, 2005.

KOENKER, R; D'OREY, V. Computing regression quantiles. **Applied statistics**. v. 36, p. 383 – 393. 1987.

KOENKER, R.; BASSETT, G. Regression Quantiles. **Econometrica**, v. 46, p. 33-50, 1978.

LÉDO, F.J. DA S.; BOTREL, M. DE A.; EVANGELISTA, A. R.; VIANA, M. C. M.; PEREIRA, A. V.; SOUZA SOBRINHO, F. DE; OLIVEIRA, J. S.; XAVIER, D.F.; HEINEMANN, A. B. Adaptabilidade e estabilidade de cultivares de alfafa avaliadas em Minas Gerais. **Ciência e Agrotecnologia**, v.29, p.409-414, 2005.

LIN, C.S.; BINNS, M.R. A superiority measure of cultivar performance for cultivar x location data. **Canadian Journal of Plant Science**, v.68, p.193–198, 1988.

LY, D.; HAMBLIN, M.; RABBI, I.; MELAKU, G.; BAKARE, M.; GAUCH JR., H. G.; OKECHUKWU, R.; DIXON, A. G. O; KULAKOW, P.; JANNINK, J. L. Relatedness and Genotype × Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava . **Crop Science**, v. 53, p. 1312 - 1325, 2013.

MONTGOMERY, D.C.; PECK, E.A.; VINING, G.G. **Introduction to linear regression analysis**. 5. ed. New York: John Wiley & Sons, 645p., 2012.

NASCIMENTO, A. C. C.; LIMA, J. E. de; BRAGA, M. J.; NASCIMENTO, M.; GOMES, A. P. Eficiência técnica da atividade leiteira em Minas Gerais: uma aplicação de regressão Quantílica. **R. Bras. Zootec.**, v.41, n.3, p.783-789, 2012

NASCIMENTO, M.; CRUZ, C.D.; CAMPANA, A.C.M.; TOMAZ, R.S.; SALGADO, C.C.; FERREIRA, R. de P. Alteração no método centroide de avaliação da adaptabilidade genotípica. **Pesquisa Agropecuária Brasileira**, v.44, p.263-269, 2009a.

NASCIMENTO, M.; FERREIRA, A.; CAMPANA, A.C.M.; SALGADO, C.C.; CRUZ, C.D. Multiple centroid methodology to analyze genotype adaptability. **Crop Breeding and Applied Biotechnology**, v.9, p.8-16, 2009b.

NASCIMENTO, M.; FERREIRA, A.; FERRÃO, R. G.; CAMPANA, A. C. M.; BHERING, L. L.; CRUZ, C. D.; FERRÃO, M. A. G.; FONSECA, A. F. A. da. Adaptabilidade e estabilidade via regressão não paramétrica em genótipos de café. **Pesquisa Agropecuária Brasileira**, v. 45, p. 41-48, 2010.

NASCIMENTO, M; SILVA, F. F. e; SÁFADI, T.; NASCIMENTO, A. C. C.; FERREIRA, R. de P.; CRUZ, C. D. Abordagem bayesiana para avaliação da adaptabilidade e estabilidade de genótipos de alfafa. **Pesquisa Agropecuária Brasileira**, v. 46, p. 26-32, 2011.

PEREIRA, H.S.; MELO, L.C.; PELOSO, M.J.D.; FARIA, L.C. de; COSTA, J. G. C. da, DÍAZ, J. L. C.; RAVA, C. A.; WENDLAND, A. Comparação de métodos de análise de adaptabilidade e estabilidade fenotípica em feijoeiro-comum. **Pesquisa agropecuária brasileira**, v.44, p.374-383, 2009.

PLAISTED, R. L.; PETERSON, L. C. A technique for evaluating the ability of selections to yield consistently in different locations and seasons. **Amer. Pot. J.**, Orono, v. 36, p. 381-385, 1959.

RESENDE, M. D. V. de; STURION, J. A. HIGA, A. R. Comparação entre métodos de avaliação da estabilidade fenotípica e adaptabilidade aplicados a dados de *Eucalyptus cloesiana* (F. MUELL.). **Boletim de Pesquisa Florestal**, Colombo, v. 1, n.42, p. 3-34, 2001.

ROCHA, R.B.; ABAD, J.I.M.; ARAUJO, E.F.; CRUZ, C.D. Avaliação do método centróide para estudo de adaptabilidade ao ambiente de clones de *Eucalyptus grandis*. **Ciência Florestal**, v. 15, p. 255-266, 2005.

SILVA, E. N da; PORTO JÚNIOR, S. S. Sistema financeiro e crescimento econômico: Uma aplicação de regressão Quantílica. **Econ. Aplic.**, v. 10, p. 425-442, 2006.

TAI, G. C. C. Genotypic stability analysis and its application to potato regional trials. **Crop Science**, Madison, v. 11, p. 184-190, 1971.

THEIL, H. A rank-invariant method of linear and polynomial regression analysis. **Indagationes Mathematicae**, v.12, p.85-91, 1950.

YAN, W.; TINKER, N. A. Biplot analysis of multi-environment trial data: Principles and applications. **Canadian journal of plant Science**, v. 86, p. 623 – 645, 2006.

## CAPÍTULO 1

### **Adaptabilidade e estabilidade para genótipos de alfafa via regressão quantílica**

#### **Resumo:**

No melhoramento genético de plantas, quando o objetivo é selecionar ou recomendar genótipos para o plantio, o estudo pormenorizado da interação entre genótipo x ambiente é de extrema importância. Para tanto, a literatura apresenta inúmeros métodos baseados em modelos de regressão que medem a resposta de cada fenótipo frente às variações ambientais por meio de uma relação funcional. Embora úteis, nenhum destes métodos trata a modelagem de maneira adequada quando o fenótipo de interesse apresenta um comportamento assimétrico ou quando apresenta *outliers*. Nessas situações, ao utilizar as metodologias usuais de adaptabilidade e estabilidade, baseadas em regressão, a relação funcional entre o fenótipo e a variação ambiental é explicada por meio de um comportamento médio, porém para estas situações o efeito do parâmetro de adaptabilidade pode ser sub ou superestimado levando a uma possível recomendação errônea. Desta forma, este trabalho tem por objetivo propor e avaliar uma metodologia de análise de adaptabilidade e estabilidade fenotípica de genótipos de alfafa baseada em regressão quantílica. Para tanto, utilizou-se um conjunto de dados provenientes de um experimento sobre produção de matéria seca de 92 genótipos de alfafa (*Medicago sativa*) avaliados em 20 ambientes. Foram simulados valores fenotípicos, com distribuições simétrica, simétrica com *outliers*, assimétrica à direita, assimétrica à direita com *outliers*, assimétrica à esquerda, assimétrica à esquerda com *outliers* para avaliação da metodologia proposta. Diante dos resultados apresentados e visto a importância econômica do lançamento de um novo cultivar, percebe-se que o tratamento adequado para fenótipos que apresentam assimetria e *outliers* é de extrema importância. O método proposto e avaliado pelos autores obteve bons resultados e desta forma, recomenda-se o uso da RQ para a adaptabilidade e estabilidade fenotípica.

**Palavras-chave:** *Medicago sativa*, interação genótipo x ambientes, melhoramento vegetal.

## 1. Introdução

No melhoramento genético de plantas, quando o objetivo é selecionar ou recomendar genótipos para o plantio, o estudo da interação entre genótipo x ambiente é de extrema importância. Entretanto, tal estudo não fornece informações pormenorizadas sobre o comportamento de cada cultivar diante das variações ambientais (CRUZ et al., 2012). Desta forma, tornam-se necessárias as análises de adaptabilidade e de estabilidade para a identificação e recomendação de materiais superiores em diferentes ambientes.

Na literatura encontram-se inúmeros métodos para realização da análise de adaptabilidade e estabilidade. Como por exemplo, as metodologias baseadas em regressão linear múltipla (CRUZ et al., 1989), regressão linear simples (FINLAY e WILKINSON, 1963; EBERHART e RUSSELL, 1966), centroides múltiplos e centroide ampliado (NASCIMENTO et al., 2009a, 2009b) e, as análises não paramétricas (LIN e BINNS, 1988; CARNEIRO, 1998).

Embora úteis, nenhum destes métodos leva em consideração a presença de fenótipos não normais, ou seja, distribuições de valores fenótipos assimétricos ou com caudas pesadas. Desta forma, caso haja a presença desse tipo de fenótipos, os métodos podem sofrer a influência de modo que a recomendação pode ser errônea. Por exemplo, trabalhos que utilizaram metodologias baseadas em regressão, como os de Vasconcelos et al. (2008) e Miranda et al. (2009), podem ter o efeito do parâmetro de adaptabilidade sub ou superestimado.

Além desses, outras metodologias tais como as propostas por Lin e Binns (1988) e Carneiro (1998), podem também ocasionar resultados impróprios, inflacionando o valor do parâmetro e com isso provocando a incorreta classificação do genótipo quanto à adaptabilidade e estabilidade. Ademais, quando se consideram as metodologias dos centroides múltiplos e centroide ampliado (NASCIMENTO et al., 2009a, 2009b), baseadas em componentes principais, a presença de pontos discrepantes influencia diretamente a configuração gráfica obtida por meio dos escores dos genótipos. Na prática, a presença de *outliers* e assimetria podem proporcionar estimativas inadequadas, que não refletem a verdadeira relação existente entre a variação ambiental e a resposta fenotípica.

Visando contornar o problema de distribuições fenotípicas com caudas pesadas, Nascimento et al. (2010) propuseram a obtenção dos estimadores utilizando-se

medianas, de forma a tornar as estimativas dos parâmetros menos sensíveis a pontos extremos que os estimadores clássicos, obtidos pelo método dos mínimos quadrados ordinários. Entretanto, tal metodologia é indicada apenas para um caso particular de distribuições de valores fenotípicos que não apresentam normalidade.

Uma solução interessante para tratar este problema de maneira unificada, isto é, a presença de pontos discrepantes ou assimetria, é a utilização de regressão quantílica (RQ). Tal metodologia, diferentemente dos métodos de regressão usuais, que utilizam a média condicional para explicar a relação funcional entre a variação ambiental e a resposta fenotípica, faz uso de funções quantílicas condicionais. Desta forma, a RQ possibilita escolher o quantil que melhor representa a relação funcional de interesse com o intuito de contemplar naturalmente a mencionada falta de normalidade.

A RQ tem sido utilizada em estudos na área de zootecnia, medicina humana e economia. Como por exemplo, Silva e Porto Júnior (2006) usaram a RQ para fornecer uma visão mais detalhada dos impactos gerados pelo sistema financeiro na distribuição condicional do crescimento econômico. Beyerlein et al. (2011) utilizaram RQ em análises GWAS (*Genome-Wide Association Study*) na área de medicina humana e enfatizaram as vantagens estatísticas e biológicas ao se estimar efeitos de marcadores em diferentes quantis das distribuições dos fenótipos. Além desses, Nascimento et al. (2012) utilizaram a RQ para investigar os determinantes da eficiência de produtores de leite de MG, para diferentes níveis de eficiência.

Diante do exposto, este trabalho teve por objetivo propor e avaliar uma metodologia de análise de adaptabilidade e estabilidade baseada em regressão quantílica. O método proposto foi utilizado para analisar um conjunto de 92 genótipos de alfafa (*Medicago sativa*), submetidos a 20 cortes. Para avaliar a metodologia, foram simulados 100 genótipos em seis situações distintas, e assim os valores das estimativas obtidas pelo método proposto foram então comparados com aqueles advindos dos métodos de Eberhart e Russell (1966) e regressão não paramétrica (NASCIMENTO et al. 2010).

## **2. Material e Métodos**

### **2.1. RQ em estudos de adaptabilidade e estabilidade**

O modelo estatístico para avaliação da adaptabilidade e estabilidade considerando  $a$  ambientes,  $g$  genótipos e  $r$  repetições pode ser definido da seguinte forma:

$$Y_{ij} = \beta_{0i}(\tau) + \beta_{1i}(\tau)I_j + e_i(\tau) \quad (1)$$

em que  $\beta_{0i}(\tau)$  é a constante da regressão;  $\beta_{1i}(\tau)$  é o coeficiente da regressão;  $e_i(\tau)$  são os erros aleatórios independentes e identicamente distribuídos com quantil de ordem  $\tau$

igual a zero;  $I_j$  é o índice ambiental codificado  $\left( I_j = \frac{\sum_i Y_{ij}}{g} - \frac{\sum_i \sum_j Y_{ij}}{ga} \right)$  e  $\tau$  refere-se ao quantil estimado ( $\tau \in [0,1]$ ).

As estimativas de  $I_j$  são de grande importância, pois indicam a qualidade do ambiente, para os quais valores negativos identificam ambientes desfavoráveis e valores positivos, ambientes favoráveis.

Na metodologia proposta, o conceito de adaptabilidade refere-se à capacidade de os genótipos responderem de maneira vantajosa ao estímulo do ambiente. Desta forma, quando  $\beta_{1i}(\tau) = 1$  os genótipos serão classificados como de adaptabilidade ampla ou geral;  $\beta_{1i}(\tau) > 1$ , genótipos com adaptabilidade específica a ambientes favoráveis;  $\beta_{1i}(\tau) < 1$ , genótipos com adaptabilidade específica a ambientes desfavoráveis.

Para a estimação do parâmetro de estabilidade, utilizou-se o coeficiente de determinação da RQ, dado por:

$$R^2(\tau) = \frac{Var(\hat{y}_{ij})}{Var(y_{ij})} \quad (2)$$

em que  $Var(\hat{y}_{ij})$  é a variância dos valores estimados e  $Var(y_{ij})$  é a variância dos valores observados.

O conceito de estabilidade refere-se à capacidade de os genótipos mostrarem um comportamento previsível em função do estímulo do ambiente. E, para medir a estabilidade, foi utilizado o parâmetro  $R^2(\tau)$ , que neste trabalho representa a proporção da variabilidade dos valores da variável dependente explicada pela variável

independente. A partir daí, pode-se classificar os genótipos de duas maneiras, genótipos com estabilidade ou previsibilidade alta, quando  $R^2(\tau)$  é maior que 90% ou genótipos com estabilidade ou previsibilidade baixa, quando  $R^2(\tau)$  é menor que 90%. No trabalho de Nascimento et al. (2010) considerou-se genótipos de previsibilidade alta aqueles em que o  $R^2$  fosse maior que 70%, entretanto visando obter um resultado mais preciso os autores optaram por 90%.

Para a estimação dos coeficientes da RQ,  $\hat{\beta}_{0i}(\tau)$  e  $\hat{\beta}_{1i}(\tau)$ , é necessário minimizar a soma dos resíduos ponderados  $y_i - \hat{y}_i$ , de maneira que resíduos com valores positivos recebem peso  $\tau$  e resíduos negativos recebem peso  $1-\tau$ . Portanto, devem-se obter os estimadores  $\hat{\beta}_{0i}(\tau)$  e  $\hat{\beta}_{1i}(\tau)$  que minimizem a seguinte equação:

$$\sum_{i=1}^n d_{\tau}(y_i, \hat{y}_i) = \tau \sum_{y_i \geq \beta_{0i}(\tau) + X_i \beta_{1i}(\tau)} |y_i - \beta_{0i}(\tau) - X_i \beta_{1i}(\tau)| + (1-\tau) \sum_{y_i < \beta_{0i}(\tau) + X_i \beta_{1i}(\tau)} |y_i - \beta_{0i}(\tau) - X_i \beta_{1i}(\tau)| \quad (3)$$

em que  $d_{\tau}$  é a distância entre  $y_i$  e  $\hat{y}_i$ .

A minimização da equação 2, permite estimar os coeficientes da regressão ( $\hat{\beta}_0(\tau), \hat{\beta}_1(\tau)$ ), o que é feito com base em algoritmos de problemas de programação linear, conforme descrito em Hao e Naiman (2007). O algoritmo citado está implementado no software R no pacote quantreg (Koenker, 2012).

Após a estimação dos parâmetros, as hipóteses de interesse,  $H_0 : \beta_{1i}(\tau) = 1$  versus  $H_1 : \beta_{1i}(\tau) \neq 1$ , são avaliadas pelo teste t, cuja estatística é dada por

$$t = \frac{\hat{\beta}_{1i}(\tau) - 1}{\sqrt{\hat{V}(\hat{\beta}_{1i}(\tau))}}. \text{ Esta estatística está associada ao número de graus de liberdade do}$$

resíduo da análise de variância conjunta e ao nível de significância  $\alpha$ . Em estudos de adaptabilidade e estabilidade, a variância de  $\beta_1(\tau)$  é dada por  $\sigma_e^2 = QMR/r$ , onde r é o número de repetições que deram origem às médias submetidas à análise de adaptabilidade e estabilidade e  $QMR = \hat{\sigma}^2$  é o quadrado médio do resíduo da análise de variância conjunta.

## 2.2. Dados simulados

Para avaliação da metodologia proposta, foram simulados 100 genótipos considerando o seguinte modelo:

$$y_{ij} = \beta_{0i} + \beta_{1i}I_j + e_{ij}, \quad (4)$$

em que:  $y_{ij}$  é a média de genótipo  $i$  no ambiente  $j$ ;  $\beta_{0i}$  é a constante da regressão referente ao  $i$ -ésimo genótipo;  $\beta_{1i}$  é o coeficiente de regressão, que mede a resposta do  $i$ -ésimo genótipo à variação do ambiente;  $I_j$  é o índice ambiental e  $e_i$  são os erros aleatórios gerados independentemente e identicamente distribuídos (i.i.d.) com distribuição de probabilidade de interesse. Considerou-se que os genótipos foram avaliados em 20 ambientes. Deve ser ressaltado que a simulação deste conjunto foi realizada levando em consideração os valores do índice ambiental ( $I_i$ ) e  $\beta_0$ , que representa a média geral do experimento, obtidos a partir do conjunto de dados avaliados no estudo de Nascimento et al. (2011) e  $\beta_1$ .

Foram consideradas 6 situações distintas. Especificamente, tais situações são caracterizadas pela simulação de valores fenotípicos,  $y_i$ , com distribuições simétrica, simétrica com *outliers*, assimétrica à direita, assimétrica à direita com *outliers*, assimétrica à esquerda, assimétrica à esquerda com *outliers*.

A inserção de assimetria à direita e à esquerda foi realizada somando-se e subtraindo-se, aos valores fenotípicos, resíduos amostrados de uma distribuição exponencial com parâmetro igual a  $\sqrt{\frac{r}{\text{QMR}}}$ , onde  $r$  é o número de repetições ( $r = 2$ ) e QMR (QMR = 55851) (NASCIMENTO et al., 2011) é o quadrado médio do resíduo da

análise conjunta, ou seja,  $e_i \sim \text{Exp}\left(\sqrt{\frac{r}{\text{QMR}}}\right)$ . Assim, os valores fenotípicos são dados por:

$$y_{iad} = y_i + \exp\left(\sqrt{\frac{r}{\text{QMR}}}\right) \text{ e } y_{iae} = y_i - \exp\left(\sqrt{\frac{r}{\text{QMR}}}\right),$$

em que  $y_{iad}$  é o  $i$ -ésimo valor do fenótipo com distribuição assimétrica a direita e  $y_{iae}$  é  $i$ -ésimo valor do fenótipo com distribuição assimétrica a esquerda.

Visando contemplar situações em que a distribuição dos valores fenotípicos é simétrica simularam-se resíduos independentes e identicamente distribuídos como uma distribuição normal com média zero e variância  $\sigma_e^2$ , ou seja,  $e_i \sim N(0, \sigma_e^2)$ .

Para a simulação dos valores fenotípicos com presença de *outliers* considerou-se como ponto de corte para a medida de influência DFBETA, o valor de  $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{20}} = 0,44$ , ou seja, o conjunto de observações simuladas que apresente valor de DFBETA superior a 0,44 é considerado com a presença de *outlier* (MONTGOMERY et al., 2012). Para simular tais valores fenotípicos, utilizou-se como valor observado 1,5 vezes à amplitude interquartil acrescido de uma constante escolhida de maneira adequada para que o conjunto de dados apresentasse *outlier* (MONTGOMERY e RUNGER, 2009).

### 2.3. Comparação de métodos de adaptabilidade e estabilidade

A fim de avaliar o desempenho da regressão quantílica na estimação dos coeficientes de regressão ( $\hat{\beta}_{1i}$ ), o processo de simulação apresentado na seção anterior foi repetido 100 vezes permitindo assim, os cálculos dos Erros Quadráticos Médios (EQM) associados a cada estimador e a porcentagem de acerto do teste, uma vez que se testaram valores de  $\beta_1 = 1$ . Os EQM foram calculados da seguinte

forma:  $EQM(\beta_{1i})_i = \frac{1}{100} \sum_{k=1}^{100} (\hat{\beta}_{1ik} - 1)^2$ , em que  $\hat{\beta}_{1ik}$  foi o coeficiente de regressão estimado do genótipo  $i$  na  $k$ -ésima repetição e  $EQM(\beta_{0i})_i = \frac{1}{100} \sum_{k=1}^{100} (\hat{\beta}_{0ik} - 1176)^2$ , em que  $\hat{\beta}_{0ik}$  é a constante da regressão estimada do genótipo  $i$  na  $k$ -ésima repetição.

Os valores das estimativas obtidas pelo método proposto foram então comparados com aqueles advindos dos métodos de Eberhart e Russell (1966) e regressão não paramétrica (NASCIMENTO et al. 2010). Foram utilizadas as funções quantílicas  $\tau = 0.25$ ,  $\tau = 0.50$  e  $\tau = 0.75$  para a estimação dos efeitos do parâmetro de adaptabilidade. Vale ressaltar que os valores de  $\tau$  foram escolhidos arbitrariamente pelos autores, o que não implica que estes são os que melhor descrevem o conjunto de dados.

Após obtenção da estimativa do parâmetro de adaptabilidade, através dos métodos avaliados neste estudo, (Eberhart e Russell (1966), regressão não paramétrica e QR (0.25, 0.50 e 0.75)), foi aplicado o teste  $t$  para verificar se existe diferença significativa na hipótese avaliada,  $H_0 : \beta_{1i} = 1$ . Posteriormente, foram calculadas as porcentagens de acerto considerando que os dados foram simulados de acordo com o modelo (4).

#### 2.4. Dados Reais

Foram utilizados dados de um experimento conduzido pela Embrapa Pecuária Sudeste, para o desenvolvimento de genótipos de alfafa adaptados a diferentes ecossistemas brasileiros. Neste experimento, que utilizou o delineamento de blocos ao acaso com duas repetições, foi avaliada a produção de massa de matéria seca de 92 genótipos de alfafa (*Medicago sativa*), submetidos a 20 cortes no período de novembro de 2004 a junho de 2006. Os cortes foram caracterizados como diferentes condições ambientais, porque foram realizados em diferentes épocas.

A avaliação da adaptabilidade e estabilidade dos genótipos foi realizada pela metodologia aqui proposta e pelas metodologias de Eberhart e Russell (1966) e de Nascimento et al. (2010). Da mesma forma que no estudo de Nascimento et al. (2010),

definiu-se como medida da influência de um ponto, a variação em módulo entre os estimadores do coeficiente de inclinação estimados pelos métodos de mínimos quadrados ( $\hat{\beta}_{ii}^*$ ) e o método de regressão não-paramétrica ( $\hat{\beta}_{ii}$ ) para cada genótipo, isto é,  $\Delta\hat{\beta}_{ii} = |\hat{\beta}_{ii}^* - \hat{\beta}_{ii}|$  em que um valor de 0,05 indica a existência de um ponto extremo.

Para verificar a assimetria dos dados, no presente trabalho foi utilizado o teste de D'Agostino (1970), cuja hipótese de nulidade a ser testada é se os dados amostrais são provenientes de uma população normalmente distribuída, ou seja, simétrica com média  $\mu$  e variância  $\sigma^2$ . Sua estatística é dada por:

$$D = \frac{T}{n^2 s},$$

em que  $s$  é o desvio padrão amostral,  $n$  é o tamanho da amostra e  $T = \sum_{i=1}^n \left(i - \frac{n+1}{2}\right) y_i$ .

Caso a amostra apresente distribuição normal, obtêm-se:

$$E(D) = \frac{(n-1)\Gamma\left(\frac{n}{2} - \frac{1}{2}\right)}{2\sqrt{2n\pi}\Gamma\left(\frac{n}{2}\right)} \approx 0,282 \text{ e } s(D) = \sqrt{\frac{12\sqrt{3} - 37 + 2\pi}{24n\pi}} \approx \frac{0,02998}{\sqrt{n}}$$

Desta forma, é possível utilizar a estatística D padronizada:

$$D^* = \frac{D - E(D)}{s(D)}$$

que apresenta distribuição normal padrão aproximada sob a hipótese nula.

## 2.5. Aspectos Computacionais

Em todo o processo de simulação e avaliação da adaptabilidade e estabilidade do método proposto utilizou-se o software livre R (R Development Core Team, 2012). O ajuste do modelo QR foi realizado utilizando-se a função *rq* do pacote *quantreg* (KOENKER, 2012). Para o ajuste do modelo da regressão não paramétrica utilizou-se a função *mblm* do pacote *mblm* (KOMSTA, 2013). A aplicação do teste de D'Agostino (1970) foi realizada utilizando o pacote *moments* (KOMSTA, 2012).

As rotinas computacionais implementadas no software R estão apresentadas no Apêndice A.

### 3. Resultados e Discussão

#### 3.1. Dados simulados

De acordo com os valores de  $EQM(\beta_1)$  observados na Tabela 1, a RQ ( $\tau = 0,50$ ) que considera a função quantílica mediana, apresentou resultados semelhantes àqueles obtidos pelo Eberhart e Russell (1966), em todas as situações avaliadas (Tabela 1). Quando são considerados fenótipos com distribuições assimétricas, verifica-se ainda, uma redução nos valores de  $EQM(\beta_1)$  quando são utilizadas as funções quantílicas RQ ( $\tau = 0,25$ ) e RQ ( $\tau = 0,75$ ), respectivamente para distribuições assimétricas à direita e à esquerda (Tabela 1).

Além disso, ao analisar as porcentagens de acerto dos dados simulados verifica-se que a RQ obteve o resultado esperado para cada distribuição estudada, sendo que quando os fenótipos são simétricos os valores de  $EQM(\beta_1)$ ,  $EQM(\beta_0)$  e a porcentagem de acerto estiveram próximos do método de Eberhart e Russell (1966).

A regressão não paramétrica, que segundo Nascimento et al. (2010) é menos influenciada por pontos extremos, obteve resultados inferiores a RQ para os casos onde a distribuição do fenótipo é assimétrica.

**Tabela 1.** Resultados dos valores de EQM e Porcentagem de acerto para as metodologias de Eberhart e Russell (1966), Regressão Quantílica e Regressão não paramétrica.

Distribuição do fenótipo	Método	$EQM(\beta_1)$	$EQM(\beta_0)$	Porcentagem de acerto
Assimetria à direita	Eberhart e Russell	0,0043	29120,59	94,86
	RQ ( $\tau = 0,25$ )	0,0019	3593,41	98,91
	RQ ( $\tau = 0,50$ )	0,0044	16697,57	94,87
	RQ ( $\tau = 0,75$ )	0,0125	61592,36	78,10
	Regressão não paramétrica	0,0048	17010,99	93,37
Assimetria à esquerda	Eberhart e Russell	0,0043	145409,25	94,69

	RQ ( $\tau = 0,25$ )	0,0125	200795,44	78,20
	RQ ( $\tau = 0,50$ )	0,0044	122715,37	94,51
	RQ ( $\tau = 0,75$ )	0,0019	88891,89	98,75
	Regressão não paramétrica	0,0048	123985,42	93,28
	Eberhart e Russell	0,0043	1384412,00	97,65
Simétrico	RQ ( $\tau = 0,25$ )	0,0082	1136971,00	92,15
	RQ ( $\tau = 0,50$ )	0,0069	1385132,00	94,23
	RQ ( $\tau = 0,75$ )	0,0080	1658201,00	92,87
	Regressão não paramétrica	0,0087	1384246,00	84,30

Quando o fenótipo de interesse apresenta *outlier* (Tabela 2), para o caso em que a distribuição é simétrica, observam-se melhores resultados para a RQ ( $\tau = 0,50$ ) e para a regressão não paramétrica. Quando se consideram fenótipos com distribuições assimétricas e presença de *outliers*, verifica-se uma redução nos valores de  $EQM(\beta_1)$  quando são utilizadas as funções quantílicas RQ ( $\tau = 0,25$ ) e RQ ( $\tau = 0,75$ ), respectivamente para distribuições assimétricas à direita e à esquerda e a regressão não paramétrica (Tabela 2).

Para a simulação a RQ obteve os melhores resultados em todas as situações descritas na Tabela 2, seguida da regressão não paramétrica.

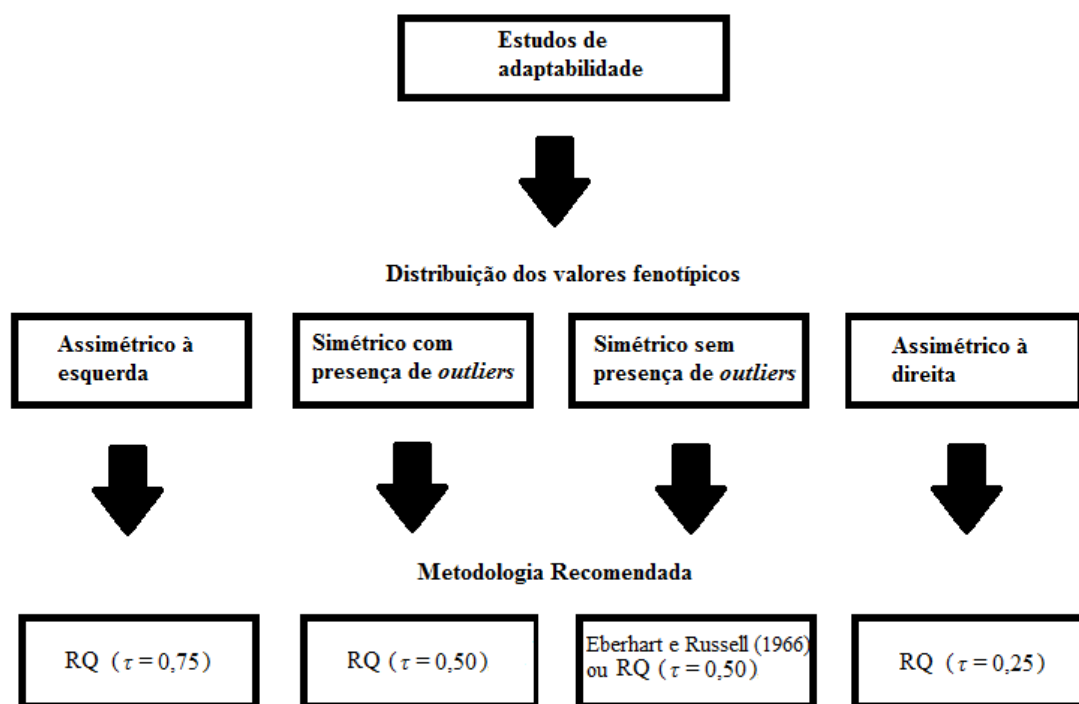
**Tabela 2.** Resultados dos valores de EQM e Porcentagem de acerto para as metodologias de Eberhart e Russell (1966), Regressão Quantílica e Regressão não paramétrica com presença de *outliers*.

Distribuição do fenótipo	Método	$EQM(\beta_1)$	$EQM(\beta_0)$	Porcentagem de acerto
Assimetria à direita com presença de <i>outliers</i>	Eberhart e Russell	0,1848	64676,76	0
	RQ ( $\tau = 0,25$ )	0,0022	3859,21	98
	RQ ( $\tau = 0,50$ )	0,0069	18143,74	89

	RQ ( $\tau = 0,75$ )	0,0974	99255,05	30
	Regressão não paramétrica	0,0060	15513,73	90
Assimetria à esquerda com presença de <i>outliers</i>	Eberhart e Russell	0.1821	4747753,00	0
	RQ ( $\tau = 0,25$ )	0,1238	4418813,00	33
	RQ ( $\tau = 0,50$ )	0,0077	5316426,00	85
	RQ ( $\tau = 0,75$ )	0,0023	5668763,00	96
	Regressão não paramétrica	0,0064	5345482,00	90
	Eberhart e Russell	0,0855	1529605,00	0
Simétrico com presença de <i>outliers</i>	RQ ( $\tau = 0,25$ )	0,0169	1095687,00	60
	RQ ( $\tau = 0,50$ )	0,0128	1350122,00	79
	RQ ( $\tau = 0,75$ )	0,05635	1788108,00	39
	Regressão não paramétrica	0,0202	1326998,00	61

Resumindo, independente do grau de assimetria do fenótipo, a RQ provém resultados similares ou melhores aos obtidos por meio dos demais métodos avaliados neste trabalho. Assim, esta metodologia torna-se uma alternativa em estudos de adaptabilidade e estabilidade.

De acordo com os resultados apresentados, sugere-se, para a escolha do método, verificar a distribuição do fenótipo de interesse, de modo que para simétricos deve-se averiguar se este possui *outlier*, se sim é utilizada a RQ ( $\tau = 0,50$ ), se não, se utiliza ou a metodologia de Eberhart e Russell (1966) ou a RQ ( $\tau = 0,50$ ). Já se o fenótipo for assimétrico utiliza-se RQ ( $\tau = 0,25$ ) para assimetria a direita e RQ ( $\tau = 0,75$ ) para assimetria à esquerda (Figura 1).



**Figura 1.** Fluxograma resumindo todos os resultados obtidos.

### 3.2. Dados Reais

Observa-se, pela Tabela 3, que houve diferenças significativas entre genótipos, o que indica a existência de variabilidade genética entre os genótipos para a produção de matéria seca. Além disso, se constatou a existência de interação genótipos  $\times$  cortes ( $P \leq 0,01$ ), desta forma tem-se a necessidade de estudo pormenorizado da resposta das cultivares frente a essas variações, por meio da análise de adaptabilidade e estabilidade, uma vez que os genótipos apresentam desempenho diferenciado frente às diferentes condições ambientais.

**Tabela 3.** Análise de variância conjunta, média geral ( $\text{kg ha}^{-1}$ ) de 92 genótipos de alfafa, para produção de matéria seca em 20 cortes em experimento conduzido de novembro de 2004 a junho de 2006

Fontes de variação	GL	Quadrados médios
Blocos	1	2 002 415.43
Cultivar	91	1 384 475.75*
Erro a	91	5 74 269.72
Corte	91	62 331 022.56*
Erro b	91	946 917.67
Interação Cultivar x Corte	1 729	60 682.46**
Erro c	1 729	55 851.26
Média ( $\text{kg ha}^{-1}$ )	1 176	

<sup>ns</sup> não significativo; \*\* e \* significativos a 5 e 1 % de probabilidade, respectivamente pelo teste F.

A Tabela 4 apresenta os 35 genótipos que obtiveram classificações de adaptabilidade discordantes de acordo com as metodologias avaliadas neste estudo. Os demais genótipos encontram-se na Tabela 6 no Apêndice B.

Dos 35 genótipos, 30 apresentaram distribuições de valores fenotípicos simétricos e 5 assimétricos à direita pelo teste D'Agostino (1970). Para os fenótipos assimétricos à direita utilizou-se a RQ no quantil 0,25 para análise da adaptabilidade, uma vez que conforme apresentado na Tabela 1 este quantil mostrou melhores resultados para fenótipos com distribuição deste tipo. Assim, a utilização do método de Eberhart e Russell (1966) pode proporcionar estimativas menos acuradas.

**Tabela 4.** Classificação da adaptabilidade dos genótipos com classificações discordantes de acordo com as três metodologias descritas no trabalho.

Genótipos	Eberhart e Russell(1966)	Regressão Quantílica ( $\tau = 0,25$ )	Regressão Quantílica ( $\tau = 0,50$ )	Regressão Não- Paramétrica	$\Delta\hat{\beta}_{1i} =$ $\left  \hat{\beta}_{1i}^{ER} - \hat{\beta}_{1i}^{NP} \right $
	Adaptab. <sup>1</sup>	Adaptab.	Adaptab.	Adaptab.	
Primavera <sup>ns</sup>	Desf.	Desf.	Geral	Geral	0,105
LE Semit 711 <sup>ns</sup>	Geral	Fav.	Geral	Fav.	0,142
Topper <sup>ns</sup>	Desf.	Geral	Desf.	Geral	0,198
LE N 3 <sup>ns</sup>	Geral	Desf.	Geral	Geral	0,017
Sequel HR <sup>ns</sup>	Geral	Geral	Geral	Fav.	0,045
Victoria SP INTA <sup>ns</sup>	Geral	Desf.	Geral	Geral	0,025
Aca 900 <sup>ns</sup>	Geral	Geral	Geral	Desf.	0,135
5 939 <sup>ns</sup>	Geral	Geral	Fav.	Geral	0,136
Medina <sup>ns</sup>	Geral	Geral	Geral	Desf.	0,093
N 910*	Geral	Geral	Geral	Desf.	0,182
Coronado*	Geral	Desf.	Desf.	Desf.	0,084
Eterna <sup>ns</sup>	Geral	Desf.	Geral	Geral	0,057
DK 193 <sup>ns</sup>	Geral	Geral	Geral	Desf.	0,092

Crioula <sup>ns</sup>	Geral	Geral	Fav.	Geral	0,050
DK 181 <sup>ns</sup>	Geral	Geral	Fav.	Geral	0,005
Activa*	Fav.	Geral	Fav.	Fav.	0,118
Aurora*	Fav.	Geral	Geral	Geral	0,177
Hunterfield <sup>ns</sup>	Geral	Geral	Geral	Fav.	0,251
P 105 <sup>ns</sup>	Geral	Fav.	Geral	Geral	0,007
WL 516*	Desf.	Geral	Geral	Geral	0,317
Tahoe <sup>ns</sup>	Geral	Geral	Fav.	Fav.	0,033
DK 167 <sup>ns</sup>	Geral	Geral	Fav.	Geral	0,054
DK 177 <sup>ns</sup>	Desf.	Geral	Geral	Geral	0,151
Maxidor <sup>ns</sup>	Geral	Geral	Fav.	Geral	0,026
Tango <sup>ns</sup>	Desf.	Desf.	Desf.	Geral	0,033
Rio Grande <sup>ns</sup>	Geral	Geral	Geral	Desf	0,165
Key II <sup>ns</sup>	Fav.	Fav.	Fav.	Geral	0,191
Lujan <sup>ns</sup>	Desf.	Geral	Geral	Geral	0,343
DK 194 <sup>ns</sup>	Geral	Geral	Desf.	Desf	0,082
P 5715 <sup>ns</sup>	Geral	Desf.	Geral	Geral	0,051
Aca 901 <sup>ns</sup>	Fav.	Geral	Fav.	Fav.	0,026
Rocio <sup>ns</sup>	Desf.	Geral	Geral	Geral	0,088
GT 13 R	Geral	Geral	Fav.	Fav.	0,051
Plus <sup>ns</sup>					
DK 187 R <sup>ns</sup>	Geral	Desf.	Geral	Desf.	0,026

Legenda: Adaptab. refere-se a Adaptabilidade, Desf. refere-se a ambientes desfavoráveis, Fav. refere-se a ambientes favoráveis. ns: não significativo a 5% de probabilidade pelo teste D'Agostino. \*: significativo a 5% de probabilidade pelo teste D'Agostino.

Ao analisar a presença de pontos extremos, tem-se que 23 genótipos apresentam  $\Delta\hat{\beta}_{ii} \geq 0,05$ , desta forma deve-se, segundo Nascimento et al. (2010), utilizar regressões baseadas na mediana, pois o ponto extremo terá influência reduzida na estimação dos parâmetros, o que evita a má interpretação do parâmetro de adaptabilidade. Para este caso utiliza-se a RQ ( $\tau = 0,50$ ) ou regressão não paramétrica.

Os genótipos Primavera, Eterna, DK 177, Lujan, DK 194, P 5715, Rocio e GT 13 R Plus são simétricos, possuem *outliers* e, além disso, apresentam a mesma classificação na regressão não paramétrica e RQ ( $\tau = 0,50$ ). Já os genótipos LE Semit

711 e Hunterfiled classificados como de adaptabilidade geral pelo método da RQ ( $\tau = 0,50$ ), são considerados de adaptabilidade específica a ambientes favoráveis pela regressão não paramétrica, o que indica que a presença do *outlier* está influenciando na estimação da reta na regressão não paramétrica. O mesmo acontece para os genótipos Aca 900, Medina, DK 193 e Rio Grande, uma vez que estes foram classificados como de adaptabilidade geral pelo método da RQ ( $\tau = 0,50$ ) e como adaptabilidade específica a ambientes desfavoráveis pela regressão não paramétrica.

Os genótipos LE N 3, Sequel HR, Victoria SP INTA, P 105, Tango, Aca 901 e DK 187 R são simétricos, não apresentam *outliers* e possuem a mesma classificação na metodologia de Eberhart e Russell (1966) e RQ ( $\tau = 0,50$ ). Os genótipos Crioula, DK 181, Tahoe e Maxidor são classificados como de adaptabilidade geral quando avaliados pelo método de Eberhart e Russell (1966), entretanto, quando são analisados com base na metodologia proposta neste trabalho (RQ ( $\tau = 0,50$ )), esses genótipos são classificados como de adaptabilidade específica a ambientes favoráveis, o que indica que a RQ ( $\tau = 0,50$ ) é mais sensível no processo de estimação que a metodologia proposta por Eberhart e Russell (1966).

Os genótipos Coronado, Activa, Aurora e WL 516 são classificados como assimétricos pelo teste D'Agostino (1970) e foram discrepantes quanto às classificações (Tabela 3), ao se comparar as estimativas dos parâmetros de adaptabilidade pela metodologia de Eberhart e Russell (1966) com a de RQ ( $\tau = 0,25$ ), e revelou a ocorrência de recomendação incorreta quando não é levada em consideração a assimetria. O genótipo WL 516 foi classificado como de adaptabilidade específica a ambientes desfavoráveis na metodologia proposta por Eberhart e Russell (1966) e no trabalho de Ferreira et al. (2004), entretanto neste trabalho, devido a presença de assimetria, este genótipo é classificado como de adaptabilidade geral.

O genótipo P 5715 foi classificado como de adaptabilidade geral pela metodologia proposta neste trabalho (RQ ( $\tau = 0,50$ )). Essa mesma classificação foi obtida no estudo de Vasconcelos et al. (2008) o que corrobora com os resultados encontrados.

Do mesmo modo que Nascimento et al. (2010), neste trabalho utilizou-se

$R^2 = \frac{Var(\hat{y}_{ij})}{Var(y_{ij})}$ , tanto para a Regressão não paramétrica quanto para a RQ, uma vez que

devido à falta de ortogonalidade no modelo, a soma de quadrados total não pode ser decomposta em soma do quadrado da regressão mais soma do quadrado do resíduo e isso pode acarretar em estimativas negativas para o primeiro parâmetro.

**Tabela 5.** Classificação da estabilidade dos genótipos de acordo com as três metodologias descritas no trabalho

Genótipos	Eberhart e Russell(1966)	Regressão Quantílica ( $\tau = 0,25$ )	Regressão Quantílica ( $\tau = 0,50$ )	Regressão Não- Paramétrica
	Estabil. <sup>1</sup>	Estabil.	Estabil.	Estabil.
Winter	Previsível	Previsível	Previsível	Previsível
SPS 6550	Previsível	Previsível	Previsível	Previsível
Primavera	Previsível	Imprevisível	Previsível	Previsível
LE N 1	Previsível	Previsível	Imprevisível	Previsível
Trindade 87	Previsível	Imprevisível	Previsível	Previsível
LE N 2	Previsível	Previsível	Previsível	Previsível
LE Semit 711	Previsível	Previsível	Previsível	Previsível
Topper	Imprevisível	Previsível	Imprevisível	Previsível
LE N 3	Previsível	Imprevisível	Previsível	Imprevisível
Sequel HR	Previsível	Previsível	Previsível	Previsível
LE N 4	Previsível	Previsível	Previsível	Previsível
Victoria SP	Previsível	Imprevisível	Previsível	Previsível
INTA				
Monarca SP	Previsível	Previsível	Previsível	Previsível
INTA				
Bárbara SP	Previsível	Previsível	Previsível	Imprevisível
INTA				
PRIMAVERA	Previsível	Imprevisível	Previsível	Previsível
Aca 900	Previsível	Previsível	Imprevisível	Imprevisível
5 939	Imprevisível	Imprevisível	Previsível	Imprevisível
WL 612	Previsível	Imprevisível	Imprevisível	Imprevisível
Medina	Previsível	Previsível	Previsível	Imprevisível
N 910	Previsível	Imprevisível	Imprevisível	Imprevisível

Coronado	Previsível	Previsível	Previsível	Imprevisível
Eterna	Previsível	Imprevisível	Previsível	Imprevisível
DK 193	Imprevisível	Imprevisível	Imprevisível	Imprevisível
Candombe	Previsível	Imprevisível	Imprevisível	Imprevisível
WL 414	Previsível	Imprevisível	Previsível	Imprevisível
Crioula	Previsível	Previsível	Previsível	Imprevisível
LE Semit 711	Previsível	Previsível	Previsível	Previsível
DK 181	Previsível	Imprevisível	Previsível	Previsível
5 929	Previsível	Imprevisível	Previsível	Imprevisível
Activa	Previsível	Imprevisível	Previsível	Imprevisível
Sequel 2	Previsível	Previsível	Previsível	Previsível
Califónia 60	Previsível	Imprevisível	Imprevisível	Imprevisível
Cuf 1010	Previsível	Previsível	Imprevisível	Previsível
58 N 58	Previsível	Previsível	Imprevisível	Imprevisível
Diamind	Previsível	Previsível	Previsível	Imprevisível
Aurora	Previsível	Imprevisível	Imprevisível	Imprevisível
Sundor	Previsível	Imprevisível	Previsível	Previsível
Springfield	Previsível	Previsível	Previsível	Previsível
Sutter	Previsível	Previsível	Imprevisível	Previsível
Hunterfield	Previsível	Imprevisível	Imprevisível	Previsível
P 105	Previsível	Previsível	Previsível	Previsível
Printa Patricia	Previsível	Previsível	Previsível	Previsível
Flórida 77	Previsível	Previsível	Imprevisível	Imprevisível
Siriver 2	Previsível	Imprevisível	Previsível	Previsível
WL 516	Imprevisível	Imprevisível	Imprevisível	Imprevisível
Tahoe	Previsível	Imprevisível	Previsível	Previsível
Esmeralda SP	Previsível	Imprevisível	Previsível	Previsível
INTA				
DK 167	Previsível	Imprevisível	Previsível	Previsível
DK 177	Previsível	Previsível	Previsível	Previsível
5 683	Previsível	Previsível	Previsível	Previsível
WL 414 1	Previsível	Previsível	Previsível	Imprevisível
Express	Previsível	Previsível	Previsível	Previsível

F 708	Previsível	Imprevisível	Previsível	Previsível
Perla SP INTA	Previsível	Previsível	Previsível	Imprevisível
Prointa Lujan	Previsível	Previsível	Previsível	Previsível
DK 166	Previsível	Imprevisível	Previsível	Imprevisível
Platino	Previsível	Previsível	Previsível	Previsível
Maxidor	Previsível	Previsível	Previsível	Previsível
Amerigraze	Previsível	Previsível	Previsível	Imprevisível
701				
13 R Supreme	Previsível	Previsível	Previsível	Previsível
Pecos	Previsível	Previsível	Previsível	Previsível
Califórnia 50	Previsível	Imprevisível	Imprevisível	Previsível
Maricopa	Previsível	Imprevisível	Previsível	Imprevisível
Kern	Previsível	Previsível	Previsível	Previsível
Costera SP	Previsível	Imprevisível	Imprevisível	Previsível
INTA				
F 686	Imprevisível	Previsível	Previsível	Previsível
Monarca	Previsível	Previsível	Previsível	Previsível
Patrícia	Previsível	Previsível	Previsível	Previsível
Tango	Previsível	Imprevisível	Previsível	Previsível
Bárbara	Previsível	Imprevisível	Imprevisível	Previsível
Rio Grande	Previsível	Imprevisível	Previsível	Imprevisível
Key II	Previsível	Imprevisível	Previsível	Imprevisível
Gala	Previsível	Previsível	Previsível	Imprevisível
Lujan	Imprevisível	Previsível	Previsível	Previsível
Perla	Previsível	Previsível	Previsível	Previsível
5683 L	Previsível	Previsível	Previsível	Previsível
Victoria	Previsível	Previsível	Imprevisível	Imprevisível
DK 194	Previsível	Imprevisível	Imprevisível	Imprevisível
WL 442	Previsível	Previsível	Previsível	Previsível
P 30	Previsível	Previsível	Previsível	Imprevisível
P 5715	Previsível	Imprevisível	Previsível	Previsível
Alfa 200	Previsível	Previsível	Previsível	Imprevisível
Aca 901	Previsível	Imprevisível	Previsível	Previsível

Gapp 969	Previsível	Previsível	Previsível	Previsível
Rocio	Previsível	Previsível	Previsível	Previsível
GT 13 R Plus	Previsível	Previsível	Previsível	Previsível
WL 525	Previsível	Imprevisível	Previsível	Previsível
Sequel	Previsível	Previsível	Previsível	Previsível
DK 187 R	Imprevisível	Imprevisível	Previsível	Imprevisível
Pinto	Previsível	Previsível	Previsível	Previsível
Bacana	Previsível	Previsível	Imprevisível	Previsível
Siriver	Previsível	Imprevisível	Imprevisível	Imprevisível

Legenda: Estabil. refere-se a Estabilidade.

Dentre os fenótipos que apresentaram assimetria, PRIMAVERA, N 910, WL 414, Activa e Aurora foram classificados como previsíveis pelo método de Eberhart e Russell (1966). Entretanto, quando foram analisados pela metodologia proposta neste trabalho com  $\tau = 0,25$ , esses genótipos foram classificados como imprevisíveis.

Os três restantes obtiveram a mesma classificação nas duas metodologias, Conrado e WL 414 1 foram classificados como previsíveis e WL 516 imprevisível.

#### 4. Conclusões

Diante dos resultados encontrados e visto a importância econômica do lançamento de um novo cultivar, percebe-se que o tratamento adequado para fenótipos assimétricos torna-se importante. O método proposto e avaliado pelos autores obteve bons resultados para a adaptabilidade e estabilidade fenotípica.

A utilização do método correto evita a má interpretação do parâmetro de adaptabilidade.

Os autores recomendam a utilização da Regressão Quantílica, pois é menos influenciada por *outliers* e por conjunto de dados que apresentam assimetria.

## Referências Bibliográficas

BEYERLEIN, A.; VON KRIES, R., NESS, A. R., ONG, K. K. Genetic Markers of Obesity Risk: Stronger Associations with Body Composition in Overweight Compared to Normal-Weight Children. **PLoS ONE**. v. 6, 2011.

CARNEIRO, P.C.S. **Novas metodologias de análise da adaptabilidade e estabilidade de comportamento**. 168p. Tese (Doutorado) - Universidade Federal de Viçosa, Viçosa. 1998.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 4.ed. Viçosa: UFV, v.1, 514p., 2012.

CRUZ, C. D.; TORRES, R. A.; VENCOVSKY, R. An alternative approach to the stability analysis proposed by Silva and Barreto. **Revista Brasileira de Genética**, v.12, p. 567-80., 1989.

D'AGOSTINO, R.B. Transformation to Normality of the Null Distribution of G1. **Biometrika**, v. 57, n.3, p. 679-681, 1970.

EBERHART, S. A.; RUSSELL, W. A. Stability parameters for comparing varieties. **Crop Science**, v.6, p.36-40, 1966.

FERREIRA, R. de P.; BOTREL, M. de A.; RUGGIERI, A.C.; PEREIRA, A.V.; COELHO, A.D.F.; LÉDO, F.J. da S.; CRUZ, C.D. Adaptabilidade e estabilidade de cultivares de alfafa em relação a diferentes épocas de corte. **Ciência Rural**, v.34, p.265-269, 2004.

FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, v.14, p.742-754, 1963.

HAO, L.; NAIMAN, D. Q. **Quantile regression**. Sage publications. 126p. 2007.

KOENKER R. quantreg: Quantile Regression. R package version 4.91. Disponível em:<http://CRAN.R-project.org/package=quantreg>, 2012.

KOMSTA , L. Moments, cumulants, skewness, kurtosis and related tests.

URL <http://www.r-project.org>;<http://www.komsta.net/>, 2012.

KOMSTA , L. Median-Based Linear Models.

URL <http://www.r-project.org>;<http://www.komsta.net/>, 2013.

LIN, C.S.; BINNS, M.R. A superiority measure of cultivar performance for cultivar x location data. **Canadian Journal of Plant Science**, v.68, p.193–198, 1988.

MIRANDA, G.V.; SOUZA, L.V. de; GUIMARÃES, L.J.M.; NAMORATO, H.; OLIVEIRA, L.R.; SOARES, M.O. Multivariate analyses of genotype x environment interaction of popcorn. **Pesquisa Agropecuária Brasileira**, v.44, p.45-50, 2009.

MONTGOMERY, D.C.; PECK, E.A.; VINING, G.G. **Introduction to linear regression analysis**. 5. ed. New York: John Wiley & Sons, 645p., 2012.

MONTGOMERY, D.C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 4 ed. Rio de Janeiro: LTC, 493 p. , 2009.

NASCIMENTO, M.; CRUZ, C.D.; CAMPANA, A.C.M.; TOMAZ, R.S.; SALGADO, C.C; FERREIRA, R. de P. Alteração no método centroide de avaliação da adaptabilidade genotípica. **Pesquisa Agropecuária Brasileira**, v.44, p.263-269, 2009a.

NASCIMENTO, M.; FERREIRA, A.; CAMPANA, A.C.M.; SALGADO, C.C.; CRUZ, C.D. Multiple centroid methodology to analyze genotype adaptability. **Crop Breeding and Applied Biotechnology**, v.9, p.8-16, 2009b.

NASCIMENTO, M.; FERREIRA, A.; FERRÃO, R. G.; CAMPANA, A. C. M.; BHERING, L. L.; CRUZ, C. D.; FERRÃO, M. A. G.; FONSECA, A. F. A. da.

Adaptabilidade e estabilidade via regressão não paramétrica em genótipos de café. **Pesquisa Agropecuária Brasileira**, v. 45, p. 41-48, 2010.

NASCIMENTO, M. ; SILVA, F.F. ; SÁFADI, T. ; NASCIMENTO, A.C.C. ; FERREIRA, R. P. ; CRUZ, C. D.. Abordagem bayesiana para avaliação da adaptabilidade e estabilidade em genótipos de alfafa. **Pesquisa Agropecuária Brasileira**. v. 46, p. 26-32, 2011

NASCIMENTO, A. C. C.; LIMA, J. E. de; BRAGA, M. J.; NASCIMENTO, M.; GOMES, A. P. Eficiência técnica da atividade leiteira em Minas Gerais: uma aplicação de regressão quantílica. **R. Bras. Zootec.**, v.41, n.3, p.783-789, 2012.

R DEVELOPMENT CORE TEAM. **R**: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012. Disponível em: <<http://r-project.org>>.

SILVA, E. N da; PORTO JÚNIOR, S. S. Sistema financeiro e crescimento econômico: Uma aplicação de regressão Quantílica. **Econ. Aplic.**, v. 10, p. 425-442, 2006.

VASCONCELOS, E.S. de; BARIONI JÚNIOR, W.; CRUZ, C.D.; FERREIRA, R. de P.; RASSINI, J.B.; VILELA, D. Seleção de genótipos de alfafa pela adaptabilidade e estabilidade da produção de matéria seca. **Acta Scientiarum. Agronomy**, v.30, p.339-343, 2008.

## CONSIDERAÇÕES FINAIS

Este trabalho abordou, de maneira detalhada, a utilização da Regressão Quantílica com uma proposta de sua utilização no melhoramento genético vegetal.

É importante analisar e discutir as diferenças entre o modelo de RQ e regressões usuais, pois a utilização do modelo correto para cada conjunto de dados pode identificar relações que não seriam possíveis utilizando outro modelo.

Desta forma, os autores sugerem a utilização da Regressão Quantílica quando seu conjunto de dados apresenta distribuições não normais.

## APÊNDICE A – Rotinas computacionais implementadas

As rotinas computacionais dos métodos descritos neste trabalho foram implementadas no software livre R (R Development Core Team, 2012) e estão descritas a seguir.

## 1. Simulação dos Dados:

```
#Função simula conjunto de dados para treinamento

####Argumentos da função
#n_ambiente = número de ambientes avaliados no estudo
#desv_pad =desvio padrão obtido por meio de
coeficiente de variação e da média do experimento em
estudo
#beta0 = média do experimento em estudo
#n_training = número de genótipos simulados
#g1 = graus de liberdade da análise conjunta
#QMR = quadrado médio do resíduo da análise conjunta
#rep = número de repetições
#I_ord = índice ambiental ordenado
```

### 1.1. Dados assimétricos à direita

```
###Simulação dos valores fenotípicos

simula<-function(n_ambientes,QMR,nrep,beta0,beta1,n_simul,
I_ord)
{
y_obs<-matrix(0,nrow=n_simul,ncol=n_ambientes,byrow = T)
for (i in 1:n_simul)
{
erro<-rexp(20,0.006)
y_obs[i,<]-beta0+beta1*I_ord+erro
dados<-y_obs
}
}
#####

####PACOTES UTILIZADOS: quantreg e mblm

###Defina os argumento da função de acordo com o
experimento avaliado
#número de indivíduos para o treinamento de cada classe
n_simul=100
#QMR
```

```

QMR=55851
#Número de repetições
nrep=2
#Média geral
beta0=1176
#beta1
beta1=1
#número de ambientes
n_ambientes=20
#Índice ambiental ordenado
I_ord<-c(-695,-681,-541,-508,-472,-351,-283,-237,-164,-
163,-29,33,45,89,112,154,203,864,1071,1553)
gl=1729
simula(n_ambientes,QMR,nrep,beta0,beta1,n_simul,I_ord)

nrepp=100
p_acertos<-matrix(NA,nrepp,5)
EQMbeta1r<-matrix(NA,nrepp,5)
EQMbeta0r<-matrix(NA,nrepp,5)

for(h in 1:nrepp)
{

n_simul=100
#QMR
QMR=55851
#Número de repetições
nrep=2
#Média geral
beta0=1176
#beta1
beta1=1
#número de ambientes
n_ambientes=20
#Índice ambiental ordenado
I_ord<-c(-695,-681,-541,-508,-472,-351,-283,-237,-164,-
163,-29,33,45,89,112,154,203,864,1071,1553)
gl=1729
simula(n_ambientes,QMR,nrep,beta0,beta1,n_simul,I_ord)

#####regressão MQO #Eberhart e Russell (1966)

difbeta1<-matrix(NA,n_simul,1)
difbeta0<-matrix(NA,n_simul,1)
teste<-matrix(NA,n_simul,1)
for(i in 1:n_simul)
{
  reg<-lm(dados[i,]~I_ord)
  b=reg$coefficients
  beta1<-b[2]
  difbeta1[i,1]<-(beta1-1)
}

```

```

beta0<-b[1]
difbeta0[i,1]<-(beta0-1176)
t<-(beta1-1)/sqrt((QMR/2)/sum(I_ord^2))
p_valor<-2*(pt(abs(t), gl, lower.tail=F))
if(p_valor>0.05)
{teste[i,1]<-1}
else {teste[i,1]<-0}

}
EQMbeta1<-sum(difbeta1^2)/n_simul
EQMbeta0<-sum(difbeta0^2)/n_simul
#EQMbeta0
p_acertos1<-(sum(teste)/n_simul)*100
#p_acertos1

p_acertos[h,1]<-p_acertos1
EQMbeta1r[h,1]<-EQMbeta1
EQMbeta0r[h,1]<-EQMbeta0

#####REGRESSÃO QUANTILICA TAU=0.25

difbeta11<-matrix(NA,n_simul,1)
difbeta00<-matrix(NA,n_simul,1)
testel<-matrix(NA,n_simul,1)
for(i in 1:n_simul)
{
  reg1<-rq(dados[i,]~I_ord,tau=0.25)
  b1=reg1$coefficients
  beta1<-b1[2]
  difbeta11[i,1]<-(beta1-1)
  beta0<-b1[1]
  difbeta00[i,1]<-(beta0-1176)
  t<-(beta1-1)/sqrt((QMR/2)/sum(I_ord^2))
  p_valor<-2*(pt(abs(t), gl, lower.tail=F))
  if(p_valor>0.05)
  {testel[i,1]<-1}
  else {testel[i,1]<-0}
}

EQMbeta11<-sum(difbeta11^2)/n_simul
#EQMbeta11
EQMbeta00<-sum(difbeta00^2)/n_simul
#EQMbeta00
p_acertos11<-(sum(testel)/n_simul)*100
#p_acertos11

EQMbeta1r[h,2]<-EQMbeta11
EQMbeta0r[h,2]<-EQMbeta00
p_acertos[h,2]<-p_acertos11

```

```
#####REGRESSÃO QUANTILICA TAU=0.5
```

```
difbeta111<-matrix(NA,n_simul,1)
difbeta000<-matrix(NA,n_simul,1)
testel<-matrix(NA,n_simul,1)
for(i in 1:n_simul)
{
  reg1<-rq(dados[i,]~I_ord,tau=0.5)
  b1=reg1$coefficients
  beta1<-b1[2]
  difbeta111[i,1]<-(beta1-1)
  beta0<-b1[1]
  difbeta000[i,1]<-(beta0-1176)
  t<-(beta1-1)/sqrt((QMR/2)/sum(I_ord^2))
  p_valor<-2*(pt(abs(t),gl,lower.tail=F))
  if(p_valor>0.05)
  {testel[i,1]<-1}
  else {testel[i,1]<-0}
}
```

```
EQMbeta111<-sum(difbeta111^2)/n_simul
#EQMbeta111
EQMbeta000<-sum(difbeta000^2)/n_simul
#EQMbeta000
p_acertos111<-(sum(testel)/n_simul)*100
#p_acertos111

EQMbeta1r[h,3]<-EQMbeta111
EQMbeta0r[h,3]<-EQMbeta000
p_acertos[h,3]<-p_acertos111
```

```
#####REGRESSÃO QUANTILICA TAU = 0.75
```

```
difbeta1111<-matrix(NA,n_simul,1)
difbeta0000<-matrix(NA,n_simul,1)
testel<-matrix(NA,n_simul,1)
for(i in 1:n_simul)
{
  reg1<-rq(dados[i,]~I_ord,tau=0.75)
  b1=reg1$coefficients
  beta1<-b1[2]
  difbeta1111[i,1]<-(beta1-1)
  beta0<-b1[1]
  difbeta0000[i,1]<-(beta0-1176)
  t<-(beta1-1)/sqrt((QMR/2)/sum(I_ord^2))
  p_valor<-2*(pt(abs(t),gl,lower.tail=F))
  if(p_valor>0.05)
  {testel[i,1]<-1}
```

```

    else {teste1[i,1]<-0}
  }

EQMbeta1111<-sum(difbeta1111^2)/n_simul
#EQMbeta1111
EQMbeta0000<-sum(difbeta0000^2)/n_simul
#EQMbeta0000
p_acertos1111<-(sum(teste1)/n_simul)*100
#p_acertos1111

EQMbeta1r[h,4]<-EQMbeta1111
EQMbeta0r[h,4]<-EQMbeta0000
p_acertos[h,4]<-p_acertos1111

#####REGRESSÃO NÃO PARAMETRICA

difbeta11111<-matrix(NA,n_simul,1)
difbeta00000<-matrix(NA,n_simul,1)
teste<-matrix(NA,n_simul,1)
for(i in 1:n_simul)
{

y<-dados[i,]
x<-I_ord

    reg<-mblm(y~x)
    b=reg$coefficients
    beta1<-b[2]
    difbeta11111[i,1]<-(beta1-1)
    beta0<-b[1]
    difbeta00000[i,1]<-(beta0-1176)
    t<-(beta1-1)/sqrt((QMR/2)/sum(I_ord^2))
    p_valor<-2*(pt(abs(t), gl,lower.tail=F))
    if(p_valor>0.05)
    {teste[i,1]<-1}
    else {teste[i,1]<-0}

}

EQMbeta11111<-sum(difbeta11111^2)/n_simul
EQMbeta00000<-sum(difbeta00000^2)/n_simul
p_acertos11111<-(sum(teste)/n_simul)*100

p_acertos[h,5]<-p_acertos11111
EQMbeta1r[h,5]<-EQMbeta11111
EQMbeta0r[h,5]<-EQMbeta00000

```

```

}
colMeans(p_acertos)
colMeans(EQMbeta0r)
colMeans(EQMbeta1r)

```

## 1.2. Dados assimétricos à esquerda

```

###Simulação dos valores fenotípicos

simula<-
function(n_ambientes,QMR,nrep,beta0,beta1,n_simul,
I_ord)
{
y_obs<-matrix(0,nrow=n_simul,ncol=n_ambientes,byrow =
T)
for (i in 1:n_simul)
{
erro<<-(-rexp(20,sqrt(nrep/QMR)))
y_obs1<-beta0+beta1*I_ord+erro
y_obs[i,<]<-y_obs1+(min(y_obs1)*-1)
dados<<-y_obs
}
}
#####

```

A partir daqui é idêntico à simulação à direita.

## 1.3. Dados simétricos

```

###Simula valores fenotípicos

simula<-function(n_ambientes,QMR,nrep,beta0,beta1,n_simul,
I_ord)
{
y_obs<-matrix(0,nrow=n_simul,ncol=n_ambientes,byrow = T)
for (i in 1:n_simul)
{
erro<<-rnorm(20,beta0,sqrt(QMR/2))
y_obs[i,<]<-beta0+beta1*I_ord+erro
dados<<-y_obs
}
}
#####

```

A partir daqui é idêntico à simulação à direita.

#### 1.4. Dados assimétricos à direita com presença de *outlier*

```
simula<-function(n_ambientes,beta0,beta1,I_ord)
{
  erro<-rexp(20,0.006)
  y_obs<-beta0+beta1*I_ord+erro
}

###Defina os argumentos da função de acordo com o
experimento avaliado
#Média geral
beta0=1176
#beta1
beta1=1
#número de ambientes
n_ambientes=20
#Índice ambiental ordenado
I_ord<-c(-695,-681,-541,-508,-472,-351,-283,-237,-164,-
163,-29,33,45,89,112,154,203,864,1071,1553)

n_genotipos=100

##com outliers
  cont<-0
  dados<-matrix(0,n_genotipos,n_ambientes,byrow=T)
  i<-1
  repeat
  {
    z1<-simula(n_ambientes,beta0,beta1,I_ord)
    reg<-lm(z1~I_ord)
    h=dfbeta(reg)
    if(sum(h[,2]>=abs(0.44))>0)
    {
      dados[i,]<-z1
      cont<-cont+1
      i<-i+1
    }
    if(cont==n_genotipos)
      break
  }

#####
```

A partir daqui é idêntico à simulação à direita.

## 1.5. Dados assimétricos à esquerda com presença de *outlier*

```
simula<-function(n_ambientes,beta0,beta1,I_ord)
{
  erro<<-(-rexp(20,0.006))
  y_obs1<-beta0+beta1*I_ord+(erro)
  y_obs<-y_obs1+(min(y_obs1)*-1)
  dados<<-y_obs
}

###Defina os argumentos da função de acordo com o
experimento avaliado
#QMR
QMR=55851
#Número de repetições
nrep=2
#Média geral
beta0=1176
#beta1
beta1=-1
#número de ambientes
n_ambientes=20
#Índice ambiental ordenado
I_ord<-c(-695,-681,-541,-508,-472,-351,-283,-237,-164,-
163,-29,33,45,89,112,154,203,864,1071,1553)

n_genotipos=100

##com outliers
cont<-0
dados<-matrix(0,n_genotipos,n_ambientes,byrow=T)
i<-1
repeat
{
  z1<-simula(n_ambientes,beta0,beta1,I_ord)
  z1<-(-z1+(max(z1)))
  reg<-lm(z1~I_ord)
  h=dfbeta(reg)
  if(sum(h[,2]>=abs(0.44))>0)
  {
    dados[i,]<-z1
    cont<-cont+1
    i<-i+1
  }
  if(cont==n_genotipos)
  break
}
#####
```

A partir daqui é idêntico à simulação à direita.

### 1.6. Dados simétricos com presença de *outlier*

```
simula<-function(n_ambientes,beta0,beta1,I_ord)
{
  erro<-rnorm(20,beta0,sqrt(QMR/2))
  y_obs<-beta0+beta1*I_ord+erro
}

###Defina os argumentos da função de acordo com o
experimento avaliado
#QMR
QMR=55851
#Número de repetições
nrep=2
#Média geral
beta0=1176
#beta1
beta1=1
#número de ambientes
n_ambientes=20
#Índice ambiental ordenado
I_ord<-c(-695,-681,-541,-508,-472,-351,-283,-237,-164,-
163,-29,33,45,89,112,154,203,864,1071,1553)

n_genotipos=100

##com outliers
cont<-0
dados<-matrix(0,n_genotipos,n_ambientes,byrow=T)
i<-1
repeat
{
  z1<-simula(n_ambientes,beta0,beta1,I_ord)
  z1[20]<-4500
  reg<-lm(z1~I_ord)
  h=dfbeta(reg)
  if(sum(h[,2]>=abs(0.44))>0)
  {
    dados[i,]<-z1
    cont<-cont+1
    i<-i+1
  }
  if(cont==n_genotipos)
  break
}
```

```
#####
```

A partir daqui é idêntico à simulação à direita.

## 2. Dados Reais

### 2.1. Adaptabilidade

```
#####MUDAR PARA O DIRETÓRIO ONDE SE ENCONTRA O  
CONJUNTO DE DADOS#####
```

```
#####PACOTES UTILIZADOS: quantreg e mblm
```

```
dados=scan("dados_reais.txt")  
dados1<-matrix(dados, 92, 21, byrow=T)  
dados1[1, ]  
I_ord<-c(-695, -681, -541, -508, -472, -351, -283, -237, -164, -  
163, -29, 33, 45, 89, 112, 154, 203, 864, 1071, 1553)  
I_ord  
QMR=55851  
nrep=2  
beta0=1176  
gl=1729
```

```
#####Eberhart and Russell(1966):  
teste<-matrix(NA, 92, 1)  
beta1<-matrix(NA, 92, 1)  
beta0<-matrix(NA, 92, 1)
```

```
for(i in 1:92)  
{  
  reg<-lm(dados1[i, -21]~I_ord)  
  b=reg$coefficients  
  beta1[i, 1]<-b[2]  
  beta0[i, 1]<-b[1]  
  t<-(beta1-1)/sqrt((QMR/2)/sum(I_ord^2))  
  p_valor<-2*(pt(abs(t), gl, lower.tail=F))  
  teste<-(p_valor>0.05)
```

```
}
```

```
write.table(beta1, "beta1.txt", row.names=FALSE, quote=FALSE)
```

```
#####REGRESSÃO NÃO PARAMÉTRICA

testeT<-matrix(NA,92,1)
beta1T<-matrix(NA,92,1)
beta0T<-matrix(NA,92,1)

for(i in 1:92)
{
y<-dados1[i,-21]
x<-I_ord
  reg<-mblm(y~x)
  b=reg$coefficients
  beta1T[i,1]<-b[2]
  beta0T[i,1]<-b[1]
  t<-(beta1T-1)/sqrt((QMR/2)/sum(I_ord^2))
  p_valor<-2*(pt(abs(t), gl,lower.tail=F))
  testeT<-(p_valor>0.05)

}

write.table(beta1T,"beta1T.txt",row.names=FALSE,quote=FALSE
)

#####REGRESSÃO QUANTILICA TAU=0.25

testeQ<-matrix(NA,92,1)
beta1Q<-matrix(NA,92,1)
beta0Q<-matrix(NA,92,1)

for(i in 1:92)
{
  reg1<-rq(dados1[i,-21]~I_ord,tau=0.25)
  b=reg1$coefficients
  beta1Q[i,1]<-b[2]
  beta0Q[i,1]<-b[1]
  t<-(beta1Q-1)/sqrt((QMR/2)/sum(I_ord^2))
  p_valor<-2*(pt(abs(t), gl,lower.tail=F))
  testeQ<-(p_valor>0.05)

}
```

```
write.table(beta1Q, "beta1Q0.25.txt", row.names=FALSE, quote=FALSE)
```

```
#####REGRESSÃO QUANTILICA TAU=0.50
```

```
teste0.5Q<-matrix(NA, 92, 1)  
beta10.5Q<-matrix(NA, 92, 1)  
beta00.5Q<-matrix(NA, 92, 1)
```

```
for(i in 1:92)  
{  
  reg1<-rq(dados1[i, -21]~I_ord, tau=0.50)  
  b1=reg1$coefficients  
  beta10.5Q[i, 1]<-b1[2]  
  beta00.5Q[i, 1]<-b1[1]  
  t<-(beta10.5Q-1)/sqrt((QMR/2)/sum(I_ord^2))  
  p_valor<-2*(pt(abs(t), gl, lower.tail=F))  
  teste0.5Q<-(p_valor>0.05)  
}
```

```
write.table(beta10.5Q, "beta1Q0.5.txt", row.names=FALSE, quote=FALSE)
```

```
#####REGRESSÃO QUANTILICA TAU=0.75
```

```
teste0.75Q<-matrix(NA, 92, 1)  
beta10.75Q<-matrix(NA, 92, 1)  
beta00.75Q<-matrix(NA, 92, 1)
```

```
for(i in 1:92)  
{  
  reg1<-rq(dados1[i, -21]~I_ord, tau=0.75)  
  b1=reg1$coefficients  
  beta10.75Q[i, 1]<-b1[2]  
  beta00.75Q[i, 1]<-b1[1]  
  t<-(beta10.75Q-1)/sqrt((QMR/2)/sum(I_ord^2))  
  p_valor<-2*(pt(abs(t), gl, lower.tail=F))  
  teste0.75Q<-(p_valor>0.05)  
}
```

```
write.table(beta1Q0.75,"beta1Q0.75.txt",row.names=FALSE,quote=FALSE)
```

## 2.2. Estabilidade

```
#####MUDAR PARA O DIRETÓRIO ONDE SE ENCONTRA O CONJUNTO DE DADOS#####
```

```
####PACOTES UTILIZADOS: quantreg e mblm
```

```
dados=scan("dados_reais.txt")
dados1<-matrix(dados,92,21,byrow=T)
dados1[1,]
I_ord<-c(-695,-681,-541,-508,-472,-351,-283,-237,-164,-163,-29,33,45,89,112,154,203,864,1071,1553)
I_ord
QMR=55851
nrep=2
beta0=1176
gl=1729
```

```
#####REGRESSÃO NÃO PARAMÉTRICA
```

```
testeT<-matrix(NA,92,1)
beta1T<-matrix(NA,92,1)
beta0T<-matrix(NA,92,1)
var1=matrix(NA,92,1)
var2=matrix(NA,92,1)
R2=matrix(NA,92,1)
```

```
for(i in 1:92)
{
```

```
  y<-dados1[i,-21]
```

```
  x<-I_ord
```

```
    reg<-mblm(y~x)
    b=reg$coefficients
    beta1T[i,1]<-b[2]
    beta0T[i,1]<-b[1]
```

```

        t<-(beta1T-1)/sqrt((QMR/2)/sum(I_ord^2))
        p_valor<-2*(pt(abs(t), gl,lower.tail=F))
        testeT<-(p_valor>0.05)
        var1[i,1]<-var(dados1[i,-21])
        var2[i,1]<-var(reg$fitted.values)
        R2[i,1]<-var2[i,1]/var1[i,1]

    }

#####REGRESSÃO QUANTILICA TAU = 0.25

testeQ<-matrix(NA,92,1)
beta1Q<-matrix(NA,92,1)
beta0Q<-matrix(NA,92,1)
var1=matrix(NA, 92,1)
var2=matrix(NA, 92,1)
R2=matrix(NA, 92,1)

for(i in 1:92)
{
    reg1<-rq(dados1[i,-21]~I_ord,tau=0.25)
    b=reg1$coefficients
    beta1Q[i,1]<-b[2]
    beta0Q[i,1]<-b[1]
    t<-(beta1Q-1)/sqrt((QMR/2)/sum(I_ord^2))
    p_valor<-2*(pt(abs(t),gl,lower.tail=F))
    testeQ<-(p_valor>0.05)
    var1[i,1]<-var(dados1[i,-21])
    var2[i,1]<-var(reg1$fitted.values)
    R2[i,1]<-var2[i,1]/var1[i,1]
}

#####REGRESSÃO QUANTILICA TAU=0.5

testeQ<-matrix(NA,92,1)
beta1Q<-matrix(NA,92,1)
beta0Q<-matrix(NA,92,1)
var1=matrix(NA, 92,1)
var2=matrix(NA, 92,1)
R2=matrix(NA, 92,1)

for(i in 1:92)
{

```

```

reg1<-rq(dados1[i,-21]~I_ord,tau=0.50)
b=reg1$coefficients
beta1Q[i,1]<-b[2]
beta0Q[i,1]<-b[1]
t<-(beta1Q-1)/sqrt((QMR/2)/sum(I_ord^2))
p_valor<-2*(pt(abs(t),gl,lower.tail=F))
testeQ<-(p_valor>0.05)
var1[i,1]<-var(dados1[i,-21])
var2[i,1]<-var(reg1$fitted.values)
R2[i,1]<-var2[i,1]/var1[i,1]
}

```

### 3. Teste de D'Agostino (1970)

```
#####PACOTE UTILIZADO: moments
```

```

dados=scan("dados_reais.txt")
dados1<-matrix(dados,92,21,byrow=T)
dim(dados1)

```

```

teste<-matrix(NA,92,1)
p_valor<-matrix(NA,92,1)
skewnes<-matrix(NA,92,1)

```

```

for(i in 1:92)
{

```

```

x= dados1[i,-21]

```

```

skewnes[i,1]<-skewness(x)
teste_agostino<-agostino.test(x)
teste_agostino
p<-teste_agostino$p.value
p_valor[i,1]<-p
teste<-(p_valor>0.05)

```

```

}

```

```

teste

```

APÊNDICE B – Tabela com os demais genótipos.

**Tabela 6.** Classificação da adaptabilidade dos genótipos com que possuem a mesma classificação nas três metodologias descritas no trabalho.

Genótipos	Eberhart e Russell(1966)	Regressão Quantílica ( $\tau = 0,25$ )	Regressão Quantílica ( $\tau = 0,50$ )	Regressão Não- Paramétrica	$\Delta \hat{\beta}_{li} =$ $ \hat{\beta}_{li}^{ER} - \hat{\beta}_{li}^{NP} $
	Adaptab. <sup>1</sup>	Adaptab.	Adaptab.	Adaptab.	
Winter <sup>ns</sup>	Geral	Geral	Geral	Geral	0,038
SPS 6550 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,067
LE N 1 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,029
Trindade 87 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,040
LE N 2 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,008
LE N 4 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,008
Monarca SP INTA <sup>ns</sup>	Geral	Geral	Geral	Geral	0,010
Bárbara SP INTA <sup>ns</sup>	Geral	Geral	Geral	Geral	0,091
PRIMAVERA*	Geral	Geral	Geral	Geral	0,027
WL 612 <sup>ns</sup>	Desf.	Desf.	Desf.	Desf.	0,096
Candombe <sup>ns</sup>	Desf.	Desf.	Desf.	Desf.	0,015
WL 414*	Desf.	Desf.	Desf.	Desf.	0,198
LE Semit 711 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,025
5 929 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,041
Sequel 2 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,002
Califónia 60 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,047
Cuf 1010 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,014
58 N 58 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,046
Diamind <sup>ns</sup>	Geral	Geral	Geral	Geral	0,064
Sundor <sup>ns</sup>	Fav.	Fav.	Fav.	Fav.	0,010
Springfield <sup>ns</sup>	Geral	Geral	Geral	Geral	0,030
Sutter <sup>ns</sup>	Geral	Geral	Geral	Geral	0,132

Prointa	Fav.	Fav.	Fav.	Fav.	0,013
Patricia <sup>ns</sup>					
Flórida 77 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,062
Siriver 2 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,019
Esmeralda SP	Geral	Geral	Geral	Geral	0,098
INTA <sup>ns</sup>					
5 683 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,013
WL 414 1*	Geral	Geral	Geral	Geral	0,061
Express <sup>ns</sup>	Geral	Geral	Geral	Geral	0,025
F 708 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,018
Perla SP	Geral	Geral	Geral	Geral	0,048
INTA <sup>ns</sup>					
Prointa Lujan <sup>ns</sup>	Fav.	Fav.	Fav.	Fav.	0,059
DK 166 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,080
Platino <sup>ns</sup>	Fav.	Fav.	Fav.	Fav.	0,034
Amerigraze	Geral	Geral	Geral	Geral	0,079
701 <sup>ns</sup>					
13 R Supreme <sup>ns</sup>	Geral	Geral	Geral	Geral	0,000
Pecos <sup>ns</sup>	Geral	Geral	Geral	Geral	0,037
Califórnia 50 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,019
Maricopa <sup>ns</sup>	Geral	Geral	Geral	Geral	0,094
Kern <sup>ns</sup>	Fav.	Fav.	Fav.	Fav.	0,007
Costera SP	Geral	Geral	Geral	Geral	0,005
INTA <sup>ns</sup>					
F 686 <sup>ns</sup>	Desf.	Desf.	Desf.	Desf.	0,062
Monarca <sup>ns</sup>	Geral	Geral	Geral	Geral	0,036
Patrícia <sup>ns</sup>	Geral	Geral	Geral	Geral	0,006
Bárbara <sup>ns</sup>	Desf.	Desf.	Desf.	Desf.	0,007
Gala <sup>ns</sup>	Geral	Geral	Geral	Geral	0,057
Perla <sup>ns</sup>	Geral	Geral	Geral	Geral	0,012
5683 L <sup>ns</sup>	Geral	Geral	Geral	Geral	0,042
Victoria <sup>ns</sup>	Geral	Geral	Geral	Geral	0,071
WL 442 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,010

P 30 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,046
Alfa 200 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,058
Gapp 969 <sup>ns</sup>	Geral	Geral	Geral	Geral	0,022
WL 525 <sup>ns</sup>	Desf.	Desf.	Desf.	Desf.	0,011
Sequel <sup>ns</sup>	Desf.	Desf.	Desf.	Desf.	0,022
Pinto <sup>ns</sup>	Geral	Geral	Geral	Geral	0,010
Bacana <sup>ns</sup>	Geral	Geral	Geral	Geral	0,011
Siriver <sup>ns</sup>	Geral	Geral	Geral	Geral	0,010

---

Legenda: Adaptab. refere-se a Adaptabilidade, Desf. refere-se a ambientes desfavoráveis, Fav. refere-se a ambientes favoráveis. ns: não significativo a 5% de probabilidade pelo teste D'Agostino. \*: significativo a 5% de probabilidade pelo teste D'Agostino.

APÊNDICE C – Demonstração de que a mediana minimiza a média da distancia absoluta

Seja  $F$  a função de distribuição acumulada e  $f$  a função densidade de probabilidade, sabe-se que:

$$\begin{aligned} E|Y - m| &= \int_{-\infty}^{+\infty} |y - m|f(y)dy = \int_{-\infty}^m |y - m|f(y)dy + \int_m^{+\infty} |y - m|f(y)dy = \\ &= \int_{-\infty}^m (m - y)f(y)dy + \int_m^{+\infty} (y - m)f(y)dy \end{aligned}$$

Além disso, tem-se que para se obter o mínimo de uma função é necessário encontrar a derivada parcial desta função e iguala-la a zero. Desta forma:

$$\begin{aligned} \frac{\partial}{\partial m} \int_{-\infty}^{+\infty} |y - m|f(y)dy &= \frac{\partial}{\partial m} \left[ \int_{-\infty}^m (m - y)f(y)dy + \int_m^{+\infty} (y - m)f(y)dy \right] = \\ &= \frac{\partial}{\partial m} \int_{-\infty}^m (m - y)f(y)dy + \frac{\partial}{\partial m} \int_m^{+\infty} (y - m)f(y)dy = \\ &= \int_{-\infty}^m \frac{\partial}{\partial m} (m - y)f(y)dy + \int_m^{+\infty} \frac{\partial}{\partial m} (y - m)f(y)dy = \\ &= \int_{-\infty}^m f(y)dy + \int_m^{+\infty} -f(y)dy = \int_{-\infty}^m f(y)dy - \int_m^{+\infty} f(y)dy = \\ &= F(m) - (1 - F(m)) = 2F(m) - 1 \end{aligned} \tag{1}$$

Igualando (1) a zero obtemos:

$$2F(m) - 1 = 0 \Rightarrow 2F(m) = 1 \Rightarrow F(m) = \frac{1}{2}$$

Como  $F(m) = \frac{1}{2}$ , tem-se que o valor de  $m$  que minimiza a média da distancia absoluta é a mediana.