

GUILHERME ALVES PUIATTI

**REGRESSÃO QUANTÍLICA NÃO LINEAR PARA DESCRIÇÃO DE  
DIFERENTES NÍVEIS DE ACÚMULO DE MATÉRIA SECA EM PLANTAS DE  
ALHO**

Tese 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 *Doctor Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2018

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

P979r  
2018 Puiatti, Guilherme Alves, 1988-  
Regressão quantílica não linear para descrição de diferentes  
níveis de acúmulo de matéria seca em plantas de alho /  
Guilherme Alves Puiatti. – Viçosa, MG, 2018.  
viii, 79f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Paulo Roberto Cecon.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.46-52.

1. Estatística agrícola. 2. *Allium sativum*. 3. Alho -  
Crescimento - Métodos estatísticos. I. Universidade Federal de  
Viçosa. Departamento de Estatística. Programa de  
Pós-graduação em Estatística Aplicada e Biometria. II. Título.

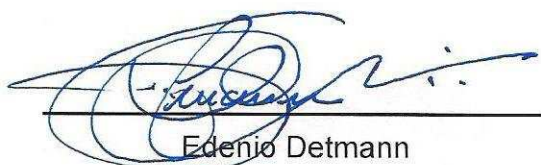
CDD 22 ed. 519.5

GUILHERME ALVES PUIATTI


**REGRESSÃO QUANTÍLICA NÃO LINEAR PARA DESCRIÇÃO DE  
DIFERENTES NÍVEIS DE ACÚMULO DE MATÉRIA SECA EM PLANTAS  
DE ALHO**

Tese 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 *Doctor Scientiae*.

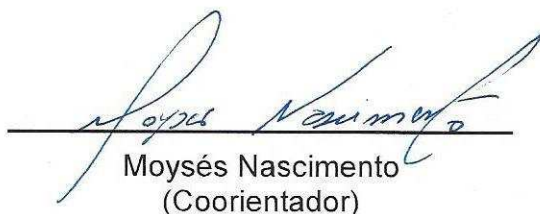
APROVADA: 19 de fevereiro de 2018.



Edenio Detmann



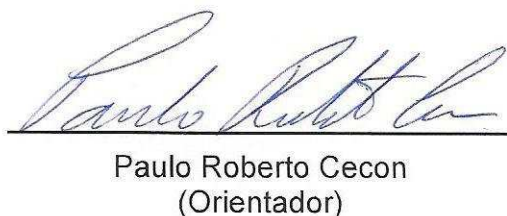
Leandro Roberto de Macêdo



Moysés Nascimento  
(Coorientador)



Ana Carolina Campana Nascimento  
(Coorientadora)



Paulo Roberto Cecon  
(Orientador)

*"A linguagem é apenas o instrumento da ciência,  
e as palavras não passam de símbolos das ideias."*

Samuel Johnson

## **AGRADECIMENTOS**

Agradeço a Deus, por me conceder forças e perseverança ao longo desta jornada.

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

À CAPES, pela concessão da bolsa de estudos.

À FAPEMIG, pelo financiamento do projeto de pesquisa.

Ao meu pai, Mário, meu maior exemplo de luta e dedicação, e minha mãe, Janice, que sempre me incentivou e acreditou em mim.

Aos meus irmãos, Gustavo e Vinícius, que estiveram ao meu lado sempre que precisei.

À minha namorada, Nádia, pela compreensão, apoio e otimismo.

Ao meu orientador e amigo, Paulo Roberto Cecon, os maiores e mais sinceros agradecimentos pela dedicação, paciência, confiança, e por todas suas palavras de sabedoria.

Aos Coorientadores Moysés Nascimento, Ana Carolina Campana Nascimento, Antônio Policarpo Souza Carneiro e Fabyano Fonseca e Silva, pelos ensinamentos e apoio ao longo do doutorado, e aos membros da banca, Leandro Roberto de Macedo e Edenio Detmann pelas sugestões para este trabalho.

Aos professores e funcionários do Departamento de Estatística da UFV, pelos conhecimentos transmitidos e pela vontade de ajudar.

Aos professores do Departamento de Fitotecnia da UFV, pela concessão dos dados do experimento.

Aos colegas do doutorado, pela convivência por estes anos, juntos dividimos nossas alegrias e aprendizados.

E a todos que contribuíram para a realização desta conquista, muito obrigado.

## **BIOGRAFIA**

GUILHERME ALVES PUIATTI, filho de Mário Puiatti e Janice Alves Pereira Puiatti, nasceu em Ponte Nova, Minas Gerais, no dia 4 de fevereiro de 1988.

Em março de 2006, ingressou no curso de Ciência da Computação da Universidade Federal de Viçosa, Viçosa – MG, graduando-se em janeiro de 2011.

Em março de 2012 iniciou o curso de Mestrado no Programa de Pós Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em 19 de fevereiro de 2014.

Em março do mesmo ano iniciou o curso de Doutorado no Programa de Pós Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, defendendo a tese em 19 de fevereiro de 2018.

## SUMÁRIO

RESUMO .....	vii
ABSTRACT .....	viii
1      INTRODUÇÃO .....	1
2      REVISÃO BIBLIOGRÁFICA .....	4
2.1    A cultura do alho.....	4
2.2    Regressão não linear .....	5
2.2.1    Método dos mínimos quadrados ordinários .....	8
2.2.1.1    Processo iterativo de Gauss-Newton .....	11
2.3    Regressão quantílica não linear .....	13
2.3.1    Algoritmo de pontos interiores para regressão quantílica não linear .....	16
2.3.1.1    Etapa dual .....	17
2.3.1.2    Etapa primal .....	18
2.3.1.3    Critério de parada.....	19
2.4    Avaliadores de qualidade de ajuste .....	19
3      MATERIAL E MÉTODOS .....	21
3.1    Descrição do experimento .....	21
3.2    Modelos de regressão não linear .....	23
3.2.1    Concordância entre modelos .....	25
3.2.2    Avaliadores de qualidade de ajuste .....	26
3.2.2.1    Quadrado médio do resíduo.....	26
3.2.2.2    Desvio médio absoluto dos resíduos.....	27
3.2.2.3    Critério de informação de Akaike .....	27
3.2.2.4    Coeficiente de determinação.....	28
3.2.2.5    Coeficiente de determinação para regressão quantílica ..	28
3.3    Distância entre acessos e modelos ajustados .....	29
3.4    Recursos computacionais.....	30
4      RESULTADOS E DISCUSSÃO .....	31
4.1    Ajuste dos modelos de regressão não linear para todos os acessos .....	31

4.2	Classificação dos acessos de acordo com sua taxa de crescimento e peso assintótico.....	34
4.3	Ajuste dos modelos de regressão não linear para cada acesso..	36
5	CONCLUSÃO .....	45
6	REFERÊNCIAS BIBLIOGRÁFICAS .....	46
	APÊNDICE .....	53
	A – Análises de concordância entre modelos .....	54
	B – Curvas dos modelos de regressão ajustados .....	55
	C – Gráficos de diagrama de caixa (boxplot) .....	71
	C – Script das análises no <i>software</i> R.....	76



## RESUMO

PUIATTI, Guilherme Alves, D.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Regressão quantílica não linear para descrição de diferentes níveis de acúmulo de matéria seca em plantas de alho.** Orientador: Paulo Roberto Cecon. Coorientadores: Moysés Nascimento, Ana Carolina Campana Nascimento, Antônio Policarpo Souza Carneiro e Fabyano Fonseca e Silva.

Análises de crescimento de plantas são importantes, pois geram informações sobre a demanda e os cuidados necessários para cada etapa de seu desenvolvimento. Modelos de regressão não linear são apropriados para descrever curvas de crescimento por apresentarem parâmetros com interpretação prática biológica. Entretanto, estes modelos apresentam informações em termos médios, e estão sujeitos a problemas no ajuste proporcionados por possíveis valores extremos ou assimetria na distribuição dos dados. A regressão quantílica pode contornar estes problemas, e ainda permite estimativas em diferentes quantis, gerando resultados mais completos e robustos. Assim, o objetivo deste trabalho foi ajustar modelos de regressão quantílica não linear para o estudo do acúmulo de matéria seca em plantas de alho ao longo do tempo e classificar cada acesso de alho de acordo com sua taxa de crescimento e peso assintótico. O modelo de regressão não linear ajustado foi o Logístico. Para tanto, utilizou-se 30 acessos de alho pertencentes ao Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV). Foram ajustados modelos para todos os acessos em três diferentes quantis, e um modelo de regressão quantílica mediana para cada acesso. Para avaliar a qualidade de ajuste dos modelos, foram utilizados o coeficiente de determinação, o quadrado médio do resíduo, o desvio médio absoluto dos resíduos, e o critério de Akaike. Os 30 acessos foram divididos de acordo com o quantil de estimativas mais próximas, sendo classificados 12 acessos como de baixo interesse para plantio, 6 de interesse intermediário e 12 como de alto interesse para plantio. A regressão quantílica não linear foi eficiente para descrição do acúmulo de matéria seca em plantas de alho ao longo do tempo, mesmo quando os dados apresentam assimetria, variâncias heterogêneas ou valores discrepantes.

## ABSTRACT

PUIATTI, Guilherme Alves, D.Sc., Universidade Federal de Viçosa, February, 2018. **Nonlinear quantile regression for the description of different levels of dry matter accumulation of garlic plants.** Adviser: Paulo Roberto Cecon. Co-advisers: Moysés Nascimento, Ana Carolina Campana Nascimento, Antônio Policarpo Souza Carneiro and Fabyano Fonseca e Silva.

Plant growth analyses are important because they generate information on the demand and the necessary cares for each of its development stages. Nonlinear regression models are appropriate to describe curves of growth, since they present parameters with biological practical interpretation. However, these models present information in terms of conditional mean, and they are subject to problems in the adjustment provided by possible outliers or asymmetry in the distribution of the data. The quantile regression can outline these problems, and it allows estimates of different quantiles, generating more complete and robust results. The objective of this work was to adjust nonlinear quantile regression models for the study of dry matter accumulation in garlic plants over time, and classify each garlic accession according to its growth tax and asymptotic weight. The nonlinear regression model fitted was the Logistic. 30 garlic accessions belonging to the Vegetable Germplasm Bank of Universidade Federal de Viçosa (BGH/UFV) were used. Models were adjusted for all of the accessions in three different quantiles, and one model of median quantile regression for each access. To evaluate the goodness of fit of the models, the coefficient of determination, the mean squared error, the absolute average residual error and the Akaike information criterion were used. The 30 accessions were divided, considering the closest quantile estimates: 12 accessions were classified as less productive, 6 of intermediate productivity and 12 as more productive. Nonlinear quantile regression was efficient for the description of dry matter accumulation in garlic plants over time, even in presence of outliers, asymmetry, or heterocedasticity in the distribution of the data.

## 1 INTRODUÇÃO

O alho (*Allium sativum* L.) é de grande importância devido aos seus valores econômico, culinário, nutricional e medicinal. As variedades desta hortaliça possuem potencial para o comércio e para a industrialização, o que tem grande influência no sistema de agricultura familiar (TRANI, 2009). O Brasil apresenta diversas variedades de alho geradas por mutações somáticas e seleções de características desejáveis para seu manejo na agricultura (Reis *et al.*, 2014), tornando necessários estudos a respeito de sua cultura e desenvolvimento. Dentre estes estudos, os que avaliam a trajetória de crescimento das plantas são de grande importância para se realizar o manejo adequado destas, pois auxiliam na preparação de técnicas para cultivo, colheita, conservação e na detecção de problemas no desenvolvimento de culturas (SOUZA & MACÊDO, 2009).

A análise de crescimento explica o desenvolvimento vegetal ao longo do tempo, gerando conhecimentos de valor prático e informações referentes ao crescimento e comportamento dos genótipos (SHARMA *et al.*, 1993). Informações a respeito de diversos cuidados, tais como manejo da adubação, podem ser identificadas através do estudo de curvas de acúmulo de matéria seca do vegetal, pois a taxa de crescimento da planta varia ao longo de seu desenvolvimento, exigindo diferentes quantidades de nutrientes em cada estágio de desenvolvimento.

Modelos de regressão não linear são apropriados para descrever curvas de crescimento. Estes modelos descrevem bem processos de crescimento, e seus parâmetros possuem interpretação biológica (MAZZINI *et al.*, 2005). As formulações destes modelos são em geral baseadas em considerações teóricas inerentes ao fenômeno que se tem interesse de modelar (MAZUCHELI e ACHCAR, 2002). Assim, é possível descrever a relação entre o acúmulo de matéria seca de plantas de alho e dias após plantio através dos parâmetros de interesse, como peso assintótico e taxa de crescimento. Estes modelos têm sido utilizados em estudos de crescimento e acúmulo de matéria seca em diferentes culturas de plantas. Pôrto *et al.* (2007) utilizaram o modelo Logístico para descrição do acúmulo de matéria seca na cultura de cebola; Maia *et al.* (2009) ajustaram modelos não lineares para

descrever o crescimento de bananeiras; Silva *et al.* (2014) utilizaram um modelo sigmoidal para analisar as taxas de acúmulo de matéria seca e micronutrientes em mandioca consorciada com bananeira irrigada; Reis *et al.* (2014) compararam diversos modelos de regressão não linear para descrever o acúmulo de massa seca de diferentes partes da planta do alho ao longo do tempo.

Entretanto, ao estudar curvas de crescimento de plantas, muitas vezes fica visível a existência de heterocedasticidade, ou seja, as variâncias entre as observações nos diferentes tempos não são sempre iguais. Ademais, pode haver o interesse em, ao ajustar um modelo para diversas subamostras, diferenciar aquelas com maiores ou menores níveis de produção. Uma alternativa para descrever a relação entre variáveis preditoras em diferentes níveis da distribuição da variável resposta pode ser a obtida pela regressão quantílica. Este método, diferente dos demais, que utilizam médias condicionais ( $E[Y|X]$ ) para obter a relação funcional entre as variáveis, utiliza quantis condicionais ( $Q[Y|X]$ ), ajustando a relação entre variável independente e quantis (percentis) da variável dependente, o que torna possível a obtenção de mais informações de localização, tais como na cauda inferior ou superior, e fornece modelos mais robustos em relação à presença de *outliers* e falta de homogeneidade (KOENKER, 2005; HAO e NAIMAN, 2007). A regressão quantílica também é aplicável a funções que são não lineares nos parâmetros.

Um benefício da regressão quantílica é a possibilidade de se ajustar modelos em diferentes quantis. Dados de crescimento na natureza normalmente apresentam diferenças em sua variação, e essas diferenças podem indicar existência de diferentes comportamentos e taxas de crescimento nas subamostras (CADE e NOON, 2003). Desta forma, a regressão quantílica não linear torna possível a identificação de diferentes níveis de crescimento entre os acessos através de diferentes quantis. Essa diferenciação é importante para melhor atender às exigências nutricionais dos diferentes acessos; qualquer falta na nutrição mineral pode causar prejuízos na produção (RESENDE *et al.*, 1999).

Alguns trabalhos estudaram curvas de crescimento de plantas utilizando modelos de regressão quantílica: Muggeo *et al.* (2013) propuseram uma estrutura de regressão quantílica utilizando *b-splines* para estimar curvas

de crescimento para alga marinha *Posidonia oceanica*; Pollice *et al.* (2014) estudaram curvas de crescimento de raízes de sorgo via regressão quantílica com *p-splines*. Sorrell *et al.* (2012) realizaram análises utilizando regressão quantílica e regressão não linear para avaliar o crescimento de três espécies de plantas de zonas úmidas em resposta a profundidade da água. No entanto, estudos que envolvem modelos de regressão quantílica não linear em acúmulo de matéria seca ainda não foram considerados na literatura.

Assim, o objetivo deste trabalho foi ajustar modelos de regressão quantílica não linear para o estudo do acúmulo de matéria seca em plantas de alho ao longo do tempo, e classificar os acessos de acordo sua taxa de crescimento e peso assintótico. Objetivou-se também comparar modelos de regressão quantílica com modelos não lineares ajustados pelo método dos mínimos quadrados ordinários.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 A cultura do alho

A planta do alho, espécie *Allium sativum* L., é herbácea, normalmente alcança 50 centímetros de altura ou mais, e possui plantas lanceoladas com limbo medindo entre 20 a 30 centímetros. Seu pseudocaule é constituído pelas bainhas das folhas as quais implantam-se em um caule pequeno e achatado. As gemas do caule formam cada uma um bulbilho, que em seu conjunto formam o bulbo. Uma característica que difere os cultivares é o número de bulbilhos que integram o bulbo, podendo ir de 5 a 56. Os bulbilhos são formados por: folha de reserva, folha de brotação e folha de proteção (SOUZA e MACÊDO, 2009).

Para que ocorra a bulbificação, a cultura do alho exige baixas temperaturas e fotoperíodos longos. Temperaturas médias de 12,8 a 23,9 °C favorecem o desenvolvimento do alho, porém existem variações entre os cultivares quanto a resposta à temperatura e ao fotoperíodo com consequente variação de resposta destas às épocas de plantio e regiões de cultivo (MUELLER *et al.*, 1990).

O alho é composto por proteínas, carboidratos, fósforo, tiamina (B1) e vitamina B6, e possui propriedades medicinais, sendo utilizado como analgésico, anti-inflamatório, antisséptico, antibacteriano, antimicótico, antiviral, anticoagulante, também protegendo o fígado, estimulando o sistema imunológico e reduzindo o colesterol. Diversos estudos a respeito do alho indicam que ele possui efeitos na prevenção de doenças cardiovasculares, reduzindo as concentrações séricas de LDL e a pressão arterial. A ação antioxidante da aliina, alicina e do ajoeno mostram o efeito do alho sobre as LDL, pois impedem a peroxidação lipídica por meio da inibição da enzima xantina-oxidase e de eicosanoides. A ação dos bioflavonóides quercetina e campferol presentes no alho eleva a capacidade total antioxidante do organismo (ANJO, 2004; MARCHIORI, 2005; TRANI, 2009).

De acordo com o IBGE (2016), a produção brasileira de alho no ano de 2015 foi de 116.764 toneladas. Os estados com a maior produção foram Minas

Gerais, com uma produção de 36.025 toneladas (30,9% da produção brasileira em 2015), Goiás com 33.241 toneladas (28,5%) e Santa Catarina com 19.011 (16,3%).

## **2.2 Regressão não linear**

Uma das tarefas usuais em análises estatísticas é a avaliação da relação entre uma variável dependente com uma ou mais variáveis independentes, e para isso modelos de regressão podem ser utilizados. Modelos de regressão linear são uma classe de modelos estatísticos que relaciona o comportamento de uma variável dependente com uma ou mais variáveis independentes e são expressos como função linear dos coeficientes de regressão (MAZUCHELI e ACHCAR, 2002). Modelos lineares simples se baseiam em pressupostos a respeito das variáveis: a relação entre variável independente e variáveis dependentes deve ser linear, e os erros devem ser independentes e ter distribuição normal, com média zero e variância homogênea (CECON *et al.*, 2012).

Modelos de regressão não linear diferem dos modelos de regressão linear principalmente por suas formulações: em geral, a formulação de modelos não lineares é baseada em considerações teóricas sobre o fenômeno que se deseja modelar (MAZUCHELI e ACHCAR, 2002). Um modelo recebe a classificação de não linear se pelo menos uma das derivadas parciais da função esperança em relação ao parâmetro resulta em uma função de parâmetros desconhecidos (PRUDENTE, 2009).

Modelos não lineares podem ser classificados como “intrinsecamente lineares” ou “intrinsecamente não lineares”. Se o modelo pode ser linearizado através de alguma transformação ou reduzido a um modelo linear através de uma reparametrização, ele é classificado como “intrinsecamente linear”. Modelos não lineares podem ser reduzidos com o objetivo de se obter estimativas iniciais para os parâmetros. Essa transformação, porém, é inconveniente porque o parâmetro normalmente perde sua interpretação intrínseca, e a estrutura e distribuição do erro são alteradas no processo. Ou seja, se os erros do modelo original satisfizerem as pressuposições usuais da

análise de regressão, os erros do novo modelo na maioria dos casos não irão satisfazê-las. Quando não se é possível obter uma reparametrização ou transformação que reduza o modelo à forma linear o modelo é considerado “intrinsecamente não linear” (MAZUCHELI e ACHCAR, 2002).

Segundo Mazucheli e Achcar (2002), modelos de regressão não linear são expressos da seguinte forma:

$$y = f(\mathbf{x}; \boldsymbol{\theta}) + \varepsilon$$

em que:

- $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  é o vetor da variável resposta (variável dependente);
- $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  é o vetor da variável explicativa (variável independente);
- $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^T$  representa o vetor de parâmetros, sendo  $p$  o número de parâmetros;
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$  é o vetor de erros aleatórios, em que, para uma amostra de tamanho suficientemente grande, assume-se  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ ;
- $f(\mathbf{x}; \boldsymbol{\theta}) = (f(x_1; \boldsymbol{\theta}), f(x_2; \boldsymbol{\theta}), \dots, f(x_n; \boldsymbol{\theta}))^T$  é uma função não linear das variáveis regressoras e dos parâmetros chamada de função esperança ou função de regressão;
- $n$  representa o número de observações.

O Método dos Mínimos Quadrados Ordinários permite ajustar um modelo com a estimativa dos parâmetros que minimiza a soma dos quadrados dos erros. Mas quando este método é aplicado em modelos não lineares, as soluções das equações normais não são facilmente obtidas analiticamente, uma vez que ambos os lados das equações ficam em função dos parâmetros desconhecidos (RATKOWSKY, 1983; GUJARATI, 2011). Assim, métodos iterativos devem ser utilizados para a obtenção das estimativas, que a partir de uma aproximação inicial gera soluções melhores a cada iteração. Dentre estes destacam-se o método de Gauss-Newton, o método Steepest-Descent (método do gradiente) e o método de Marquardt (BATES e WATTS, 1988).



Curvas de crescimento descrevem uma sequência de medidas de um indivíduo em função do tempo. Curvas de crescimento em formato sigmoidal, em que as taxas de crescimento aumentam até atingir um ponto de inflexão e depois decrescem até tenderem a zero, são adequadamente interpretadas por modelos de regressão não linear (FERNANDES *et al.*, 2015). Dessa forma, funções não lineares têm se mostrado efetivas para a análise de dados de crescimento de plantas, uma vez que essas funções sintetizam um grande número de medidas em poucos parâmetros que contém significado biológico, o que facilita a interpretação e o entendimento do fenômeno em estudo (BROWN *et al.*, 1976; DENISE e BRINKS, 1985).

Aplicações de modelos de regressão não linear são comuns em biologia, quando plantas e organismos crescem com o tempo, existindo também aplicações em Economia e Engenharia para esses tipos de modelos (REGAZZI, 2003). Os estudos do acúmulo de matéria seca e absorção de nutrientes por plantas de alho são especialmente importantes, pois os diversos nutrientes exigidos são acumulados em diferentes quantidades e etapas de crescimento e desenvolvimento da planta. Essas exigências devem ser consideradas ao realizar a adubação correta da planta, pois a falta de nutrição mineral leva a prejuízos na produção e na conservação pós-colheita dos bulbos (RESENDE *et al.*, 1999).

Dentre os modelos mais utilizados para descrição de crescimento de plantas e animais estão: Brody, Gompertz, Logístico, Richards, Mitscherlich, Weibull, e von Bertalanffy (SILVEIRA *et al.*, 2011; REIS *et al.*, 2014). Alguns destes modelos são apresentados na Tabela 1. Pôrto *et al.* (2007) utilizaram o modelo Logístico para descrever acúmulo de matéria seca na cultura de cebola. Maia *et al.* (2009), em um estudo de comparação de modelos de regressão não linear em bananeiras, fizeram uso dos modelos Brody, Gompertz, Logístico e von Bertalanffy. Reis *et al.* (2014) utilizaram diversos modelos não lineares para descrever o acúmulo de matéria seca do bulbo, da raiz e total da planta em grupos de acessos de alho.

**Tabela 1** Modelos de regressão não linear utilizados para descrever curvas de crescimento.

Modelo	Função
Brody	$y_i = \beta_1(1 - \beta_2 e^{(-\beta_3 x_i)}) + \varepsilon_i$
Gompertz	$y_i = \beta_1 e^{(-\beta_2 e^{(-\beta_3 x_i)})} + \varepsilon_i$
Logístico	$y_i = \frac{\beta_1}{1 + \beta_2 e^{(-\beta_3 x_i)}} + \varepsilon_i$
Mitscherlich	$y_i = \beta_1(1 - e^{(\beta_3 \beta_2 - \beta_3 x_i)}) + \varepsilon_i$
von Bertalanffy	$y_i = \beta_1(1 - \beta_2 e^{(-\beta_3 x_i)})^3 + \varepsilon_i$
Schnute	$y_i = \frac{\beta_1}{\left(1 + \beta_4 e^{(\beta_2 \beta_3 - x_i)}\right)^{\frac{1}{\beta_4}}} + \varepsilon_i$
Richards	$y_i = \frac{\beta_1}{\left(1 + e^{(\beta_2 - \beta_3 x_i)}\right)^{\frac{1}{\beta_4}}} + \varepsilon_i$
Meloun I	$y_i = \beta_1 - \beta_2 e^{(-\beta_3 x_i)} + \varepsilon_i$
Meloun II	$y_i = \beta_1 - e^{(-\beta_2 - \beta_3 x_i)} + \varepsilon_i$

Para estes modelos,  $\beta_1$  representa o valor máximo esperado da resposta, o valor assintótico;  $\beta_2$  é uma constante de integração;  $\beta_3$  é a taxa de crescimento da curva; e  $\beta_4$ , presente em alguns dos modelos, determina o ponto de inflexão (PRUDENTE, 2009).

### 2.2.1 Método dos mínimos quadrados ordinários

Considerando o modelo descrito por Mazucheli e Achcar (2002), a relação entre variável resposta e variável independente pode ser representada por uma equação da forma  $y = f(x; \theta) + \varepsilon$ , como descrita anteriormente. Matricialmente, o modelo é dado por:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad f(\mathbf{x}; \boldsymbol{\theta}) = \begin{bmatrix} f(x_1, \boldsymbol{\theta}) \\ f(x_2, \boldsymbol{\theta}) \\ \vdots \\ f(x_n, \boldsymbol{\theta}) \end{bmatrix} \quad \text{e} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Uma vez que a função esperança é contínua e admite derivadas de primeira e segunda ordens em relação aos parâmetros, a soma de quadrados dos erros é definida por:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(\mathbf{x}_i; \boldsymbol{\theta})]^2$$

em notação vetorial:

$$S(\boldsymbol{\theta}) = [\mathbf{y} - f(\boldsymbol{\theta})]^T [\mathbf{y} - f(\boldsymbol{\theta})] = \|\mathbf{y} - f(\boldsymbol{\theta})\|^2$$

na qual  $f(\boldsymbol{\theta}) = (f(x_1; \boldsymbol{\theta}), \dots, f(x_n; \boldsymbol{\theta}))^T$ . A soma dos quadrados dos erros  $S(\boldsymbol{\theta})$  deve ser minimizada por  $\boldsymbol{\theta}$ .  $F(\boldsymbol{\theta})$ , definida a seguir, é a matriz Jacobiana da função vetor  $f(\boldsymbol{\theta}) = (f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \dots, f_n(\boldsymbol{\theta}))^T$ . A função  $f(\boldsymbol{\theta})$  é de dimensão  $n \times 1$ , e o vetor de parâmetros  $\boldsymbol{\theta}$  de  $p \times 1$ :

$$F(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \left[ \left( \frac{\partial f_i(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \right) \right] = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}$$

O estimador de mínimos quadrados  $\hat{\boldsymbol{\theta}}$  seguirá a equação:

$$\left[ \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right]_{\boldsymbol{\theta}' = \hat{\boldsymbol{\theta}}'} = \mathbf{0}$$

$$\begin{aligned} \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} &= \frac{\partial}{\partial \boldsymbol{\theta}^T} [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]^T [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})] \\ &= -2 [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]^T \mathbf{F}(\boldsymbol{\theta}) \\ &= -2 \mathbf{F}(\boldsymbol{\theta})^T [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})] \end{aligned}$$

Dessa forma:

$$\mathbf{F}(\hat{\boldsymbol{\theta}})^T [\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}})] = \mathbf{0}$$

Sendo o sistema de equações normais para a regressão não linear dado por:

$$\begin{bmatrix} \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} \\ \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_p} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_p} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_p} \end{bmatrix} \left( \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} f_1(\hat{\boldsymbol{\theta}}) \\ f_2(\hat{\boldsymbol{\theta}}) \\ \vdots \\ f_n(\hat{\boldsymbol{\theta}}) \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Estas equações normais são não lineares em relação aos parâmetros. Assim, as soluções destas equações somente podem ser obtidas com o uso de um método iterativo.

Propriedades dos estimadores de mínimos quadrados, tais como não tendenciosidade, consistência e variância mínima dentre todos os estimadores só são válidas no caso da regressão não linear com base na teoria assintótica, isto é, para amostras de tamanho suficientemente grande (RATKOWSKY, 1983).

### 2.2.1.1 Processo iterativo de Gauss-Newton

O processo iterativo de Gauss-Newton é um dos algoritmos mais utilizados para se obter estimativas de mínimos quadrados não lineares. O método é considerado iterativo porque, a partir de um valor inicial dos parâmetros, o processo é repetido diversas vezes, gerando uma solução aproximada a cada etapa. Novas iterações são executadas até um critério de parada ser atingido. O processo é apresentado a seguir, como descrito por Mazucheli e Achcar (2002).

Supondo  $\hat{\theta}_0$ , uma aproximação inicial para  $\hat{\theta}$ , temos as expansões de Taylor das componentes de  $f(\theta)$  na vizinhança de  $\hat{\theta}_0$ :

$$f(x_i, \theta) = f(x_i, \hat{\theta}_0) + \frac{\partial f(x_i, \hat{\theta}_0)}{\partial \theta^T} (\theta - \hat{\theta}_0)$$

produzindo a aproximação linear:

$$f(\theta) \cong f(\hat{\theta}_0) + F(\hat{\theta}_0)(\theta - \hat{\theta}_0),$$

em que

$$F(\hat{\theta}_0) = \left[ \frac{\partial f(\theta)}{\partial \theta^T} \right]_{\theta = \hat{\theta}_0}.$$

Assim, para minimizar a função de soma dos quadrados dos erros deve-se minimizar a função  $S(\theta)$  associada à aproximação linear:

$$S(\theta) = \|y - f(\hat{\theta}_0) - F(\hat{\theta}_0)(\theta - \hat{\theta}_0)\|^2.$$

Considerando que  $y - f(\hat{\theta}_0) = E_0$  e  $\theta - \hat{\theta}_0 = \Delta \hat{\theta}_0$ , a partir do resultado anterior temos:

$$\begin{aligned}
S(\theta) &= \left\| E_0 - F(\hat{\theta}_0) \Delta \hat{\theta}_0 \right\|^2 \\
&= \left[ E_0 - F(\hat{\theta}_0) \Delta \hat{\theta}_0 \right]^T \left[ E_0 - F(\hat{\theta}_0) \Delta \hat{\theta}_0 \right] \\
&= E_0^T E_0 - E_0^T F(\hat{\theta}_0) \Delta \hat{\theta}_0 - \Delta \hat{\theta}_0^T F(\hat{\theta}_0)^T E_0 + \Delta \hat{\theta}_0^T F(\hat{\theta}_0)^T F(\hat{\theta}_0) \Delta \hat{\theta}_0 \\
&= E_0^T E_0 - 2 \Delta \hat{\theta}_0^T F(\hat{\theta}_0)^T E_0 + \Delta \hat{\theta}_0^T F(\hat{\theta}_0)^T F(\hat{\theta}_0) \Delta \hat{\theta}_0
\end{aligned}$$

A derivada de  $S(\theta)$  em relação a  $\Delta \hat{\theta}_0$  é:

$$\begin{aligned}
\frac{\partial S(\theta)}{\partial \Delta \hat{\theta}_0} &= 0 - 2 F(\hat{\theta}_0)^T E_0 + 2 F(\hat{\theta}_0)^T F(\hat{\theta}_0) \Delta \hat{\theta}_0 \\
&= 2 \left[ F(\hat{\theta}_0)^T F(\hat{\theta}_0) \Delta \hat{\theta}_0 - F(\hat{\theta}_0)^T E_0 \right]
\end{aligned}$$

Igualando a derivada a zero obtemos o Sistema de Equações Normais:

$$F(\hat{\theta}_0)^T F(\hat{\theta}_0) \Delta \hat{\theta}_0 = F(\hat{\theta}_0)^T E_0,$$

em que  $E_0$  desempenha o papel de variável dependente  $\mathbf{y}$  e  $F(\hat{\theta}_0)^T$  o da matriz  $\mathbf{X}$  dos modelos lineares. Se  $F(\hat{\theta}_0)$  apresenta posto coluna completo, o valor de  $\Delta \hat{\theta}_0$  que minimiza  $S(\theta)$  é:

$$\Delta \hat{\theta}_0 = \left[ F(\hat{\theta}_0)^T F(\hat{\theta}_0) \right]^{-1} F(\hat{\theta}_0)^T E_0.$$

Pode-se então definir um vetor  $\hat{\theta}_1 = \hat{\theta}_0 + \Delta \hat{\theta}_0$  de estimativas atualizadas de  $\theta$ . Seguindo este procedimento, pode-se obter utilizando  $\hat{\theta}_1$  outro conjunto de estimativas atualizadas  $\hat{\theta}_2$ , e assim por diante. Em geral, a  $k$ -ésima iteração será:

$$\begin{aligned}\hat{\theta}_{k+1} &= \hat{\theta}_k + \Delta \hat{\theta}_k \\ &= \hat{\theta}_k + \left[ F(\hat{\theta}_k)^T F(\hat{\theta}_k) \right]^{-1} F(\hat{\theta}_k)^T E_k\end{aligned}$$

em que

$$\begin{aligned}F(\hat{\theta}_k) &= \left[ \frac{\partial f(\theta)}{\partial \theta^T} \right]_{\theta=\hat{\theta}_k}; \\ E_k &= y - f(\hat{\theta}_k); \\ \hat{\theta}_k &= [\hat{\theta}_{1k} \quad \hat{\theta}_{2k} \quad \dots \quad \hat{\theta}_{pk}]^T.\end{aligned}$$

O processo se repete até a convergência, que ocorrerá quando  $|\hat{\theta}_{k+1} - \hat{\theta}_k| < \delta$ , em que  $\delta$  é algum valor fixo bem pequeno. O valor de  $S(\hat{\theta}_k)$  pode ser avaliado como garantia de sua redução durante cada iteração. Caso  $S(\hat{\theta}_{k+1})$  seja maior que  $S(\hat{\theta}_k)$  o vetor de incremento será  $\Delta \hat{\theta}_k / 2$ .

### 2.3 Regressão quantílica não linear

Segundo Koenker (2005) e Mosteller e Tukey (1977), um modelo de regressão usual ajustado pelo método dos mínimos quadrados ordinários resume toda informação das variáveis independentes observadas às suas médias. A média é uma informação resumida e incompleta de uma distribuição, da mesma forma que a regressão é uma visão limitada de um conjunto de distribuições. Uma possibilidade bem mais completa seria ajustar diversas curvas de regressão referentes a diversos quantis da distribuição.

Koenker e Bassett (1978) propuseram outra abordagem: a regressão quantílica, a qual permite calcular estimativas para quaisquer quantis da variável dependente, tais como a mediana. Isso possibilita a obtenção de mais informações de localização, como na cauda inferior ou superior, ou subpopulações específicas (HAO e NAIMAN, 2007). Dessa forma, é possível avaliar o impacto de uma covariável em toda a distribuição da variável

resposta, e não apenas na sua média. Assim, diferentes resultados de uma variável explicativa em quantis distintos podem ser interpretados como diferenças na resposta da variável dependente a mudanças nos regressores em vários pontos na distribuição condicional da variável dependente (SILVA e PORTO JÚNIOR, 2006). Além disso, estes modelos são mais robustos em relação a *outliers* e erros que não seguem distribuição normal, podendo incorporar uma possível heterocedasticidade detectada a partir da variação das estimativas dos coeficientes dos parâmetros para os diferentes quantis (BARROSO *et al.*, 2015).

A teoria da regressão quantílica evoluiu bastante desde sua proposta original, sendo estendida também para a classe dos modelos não lineares por Koenker e Park (1996). O comportamento assintótico dos estimadores dos modelos de regressão quantílica não linear se assemelha à teoria dos mínimos quadrados não lineares, de forma que a inferência para regressão quantílica não linear pode ser adaptada diretamente dos métodos destes (KOENKER, 2005). Algumas metodologias para o ajuste e implementação computacional de modelos de regressão quantílica não linear são citados a seguir.

Bouyé e Salmon (2009) desenvolveram uma aproximação para modelagem de regressão quantílica não linear utilizando funções cópula, empregadas para definir a estrutura de dependência entre as variáveis de interesse. Esta aproximação se baseia no fato de que uma distribuição conjunta pode ser dividida em duas partes, as marginais e a função de dependência (cópula).

Hunter e Lange (1999) propuseram um algoritmo iterativo para encontrar quantis amostrais, aplicável em casos de regressão quantílica não linear. Esse algoritmo recebe a terminologia de MM (*majorize-minimize*, ou majorar-minimizar) porque visa aumentar o valor da função objetivo através de uma função quadrática para então minimizar esta função quadrática.

Por sua vez, Koenker e Park (1996) consideraram um método de pontos interiores, que embora perca em eficiência para algoritmos do método simplex, se mostram como uma extensão natural para problemas não lineares. Este método é, em base, uma variação do caso de mínimos quadrados reponderados iterativamente: o sistema de equações equivale ao problema de



mínimos quadrados comuns ponderados, mas com uma dependência recursiva entre os pesos, resíduos e coeficientes.

Segundo Koenker (2005), o quantil de ordem  $\tau$  de uma variável aleatória  $Y$  pode ser definido como a inversa da função de distribuição acumulada no ponto  $\tau$ , ou seja:

$$Q_{\tau}(y) = \inf \{y : F_Y(y) \geq \tau\},$$

sendo  $F_Y(y) = P(Y \leq y)$  a função de distribuição acumulada de  $Y$ , e  $\tau \in [0,1]$ . Assim, o modelo de regressão quantílica não linear é expresso da seguinte forma:

$$Q_{\tau}(Y | X = x) = f(x; \theta(\tau))$$

em que:

- $y = (y_1, y_2, \dots, y_n)^T$  é o vetor da variável resposta;
- $x = (x_1, x_2, \dots, x_n)^T$  é o vetor da variável explicativa;
- $\theta(\tau) = (\theta_1(\tau), \theta_2(\tau), \dots, \theta_p(\tau))^T$  representa o vetor de  $p$  parâmetros no  $\tau$ -ésimo quantil de  $y$ ;
- $f(x; \theta(\tau)) = (f(x_1; \theta(\tau)), \dots, f(x_n; \theta(\tau)))^T$  é uma função não linear da variável regressora e dos parâmetros.

O uso da regressão quantílica, tanto linear quanto não linear, é mais comum na área da economia (KOENKER e HALLOCK, 2001), mas também tem sido aplicado em ecologia (CADE e NOON, 2003), sendo ainda pouco considerado para curvas de crescimento. Cao e Dean (2015) utilizaram regressão quantílica não linear baseada no 99º quantil para caracterizar o limite superior do diâmetro de árvores em densidades espaciais mais baixas; a curva de regressão quantílica descreveu de maneira mais realista a relação entre número de árvores e diâmetro máximo de árvore para uma população florestal. Schröder *et al.* (2005) utilizaram regressão quantílica não linear para estimar a

resposta de 18 espécies de plantas a diversos fatores ambientais e afirmaram que a regressão quantílica não linear foi útil para as análises, enquanto a regressão padrão utilizando a média condicional subestimou as taxas de mudança de cobertura das espécies para determinado fator.

### 2.3.1 Algoritmo de pontos interiores para regressão quantílica não linear

Koenker e Park (1996) propuseram um algoritmo de pontos interiores que tem a finalidade de computar estimativas de regressão quantílica para casos em que a função resposta é não linear nos parâmetros. O algoritmo é apresentado a seguir, conforme descrito pelos autores. Considerando o seguinte problema não linear  $L_1$ :

$$\min_{\theta \in \mathbb{R}^p} \sum |f_i(\theta)|$$

na qual, por exemplo,  $f_i(\theta) = y_i - f_0(x_i, \theta)$ . A regressão  $L_1$  (ou regressão de erros absolutos) se baseia em uma função de distância que visa minimizar a soma dos valores absolutos dos desvios (CECON *et al.*, 2012).

Uma condição necessária para que o vetor inicial de parâmetros  $\theta^*$  resolva a equação é que exista um vetor  $d \in [-1, 1]^n$  tal que:

$$J(\theta^*)^T d = 0$$

$$f(\theta^*)^T d = \sum |f_i(\theta^*)|$$

em que  $f(\theta) = (f_i(\theta))$  e  $J(\theta) = (\partial f_i(\theta) / \partial \theta_j)$ .

Uma aproximação para a solução do problema é resolver uma sucessão de problemas linearizados  $L_1$  minimizando  $\sum |f_i(\theta) - J_i(\theta)^T \delta|$ , escolhendo a cada iteração um passo de largura  $\lambda$ , procurando nas direções resultantes  $\delta$ . Surgem duas dificuldades nesta aproximação: primeiramente, é necessário resolver um problema  $L_1$  a cada iteração; e segundo, as direções

resultantes podem ser inferiores a direções determinadas pelas soluções incompletas para sequência de problemas linearizados. A informação contida nos passos iniciais da iteração dual parece mais indicada para resolver o problema não linear, uma vez que carregam a informação relevante local do valor atual de  $\theta$ .

Seja  $\theta$  o valor do parâmetro na atual iteração, e o problema dual:

$$\max \left\{ f^T d \mid d \in [-1, 1]^n, J^T d = 0 \right\}.$$

Se o modelo fosse linear, de forma que  $f(s) = f(\theta) - K(s - \theta)$  para uma matriz  $K$ , então uma solução seria possível ao aplicar o algoritmo de Meketon (1986) para encontrar  $d^*$  que resolva o problema dual, computando:

$$\delta^* = \left( K^T D^2 K \right)^{-1} K^T D^2 f,$$

em que  $D = \text{diag} \left( \min \{1 - d_i^*, 1 + d_i^*\} \right)$  e  $\theta^* = \theta + \delta^*$ . Quando  $f$  é não linear não é conveniente resolver totalmente o problema dual a cada interação. O algoritmo, então, segue duas etapas duais entre sucessivas atualizações de  $f$  e  $J$ .

### 2.3.1.1 Etapa dual

Para qualquer  $d$  viável no interior do conjunto de restrição definido acima, este é refinado seguindo o algoritmo de Meketon (1986). Sejam:

$$\begin{aligned} D &= \text{diag} \left( \min \{1 - d_i, 1 + d_i\} \right), \\ s &= D^2 \left( I - J \left( J^T D^2 J \right)^{-1} J^T D^2 \right) f, \\ d &\leftarrow d + (\eta / \alpha) s, \end{aligned}$$

em que  $\alpha = \max \left\{ \varepsilon, \max \left\{ s_i / (1 - d_i), -s_i / (1 + d_i) \right\} \right\}$ ;  $\eta \in (0, 1)$  é a constante escolhida para assegurar a viabilidade, que será  $\eta = 0,97$  conforme sugerido por Meketon (1986); e  $\varepsilon > 0$  é um parâmetro de tolerância. A iteração segue, com os novos valores de  $D$ ,  $s$  e  $d$ . Agora  $f$  e  $J$  dependem do valor atual de  $\theta$ , então o processo é incorporado em uma sequência de iterações primais nas quais  $f$  e  $J$  são atualizados.

### 2.3.1.2 Etapa primal

O passo anterior fornece a direção primal  $\delta = (J^T D^2 J)^{-1} J^T D^2 f$ . A atualização leva a:

$$\theta \leftarrow \theta + \lambda^* \delta$$

em que  $\lambda^* = \arg \min_{\lambda} \sum |f_i(\theta + \lambda \delta)|$ . São então atualizados os valores de  $f$  e  $J$ .

Antes de retornar à etapa dual,  $d$  deve ser ajustado para garantir que este seja viável para o novo valor de  $J$ . Isto é realizado projetando o  $d$  atual no espaço nulo do novo  $J$ ,  $\hat{d} = \left( I_n - J (J^T J)^{-1} J^T \right) d$  e então reduzindo este para assegurar que esteja em  $[-1, 1]^n$ , para que:

$$d \leftarrow \hat{d} / \left( \max_i \left\{ |\hat{d}_i| \right\} + \varepsilon \right),$$

sendo  $\varepsilon > 0$  um parâmetro de tolerância.

### 2.3.1.3 Critério de parada

A execução do algoritmo termina quando uma nova iteração não melhora a função objetivo em mais que uma tolerância especificada, sendo o valor igual a  $10^{-7}$  na implementação de Koenker e Park (1996).

## 2.4 Avaliadores de qualidade de ajuste

Avaliadores de qualidade de ajuste são ferramentas estatísticas que permitem quantificar o quão bem o modelo ajustado descreve os dados, possibilitando a comparação entre diferentes modelos para determinar o mais indicado (SILVEIRA *et al.*, 2011). Existem diversas medidas propostas na literatura. Segundo Motulsky e Christopoulos (2003), o modelo desejável é o mais parcimonioso, envolvendo o menor número de parâmetros que expliquem adequadamente o comportamento da variável resposta.

Alguns dos avaliadores mais utilizados são o coeficiente de determinação, o coeficiente de determinação ajustado e o desvio padrão residual. Existem dúvidas a respeito do uso do coeficiente de determinação ( $R^2$ ) para regressão não linear, pois este não possui significado prático neste caso, e a soma de quadrados total não necessariamente será igual soma de quadrados da regressão mais a soma de quadrados dos erros (RATKOWSKY, 1983). Souza (1998) afirmou que a qualidade do ajustamento pode ser medida pelo quadrado do coeficiente de correlação entre os valores observados e preditos também em modelos de regressão não linear, sendo uma medida útil para comparação de modelos, como feito nos trabalhos de Maia *et al.* (2009) e Reis *et al.* (2014). O quadrado médio do resíduo (QMR) e o desvio médio absoluto (DMA) também são medidas utilizadas no caso de comparação de curvas de crescimento, como nos trabalho de Sarmento *et al.* (2006), Maia *et al.* (2009) e Reis *et al.* (2014). O critério de informação de Akaike também é uma estatística muito utilizada para comparar a qualidade de ajuste de modelos, e é baseada no máximo da função de verossimilhança, considerando o número de parâmetros do modelo em estudo. Este critério foi utilizado por Sousa *et al.* (2012) e Silveira *et al.* (2011).

Koenker e Machado (1999) sugerem avaliadores de qualidade de ajuste específicos para regressão quantílica. Entre os avaliadores propostos, estão uma medida análoga ao  $R^2$  convencional, e também processos baseados em Testes de Razão de Verossimilhanças, de Wald e Multiplicadores de Lagrange. Estes testes medem a qualidade dos modelos para um quantil em particular e não para distribuição inteira, sendo uma medida para a qualidade de ajuste local para o quantil em questão.

### **3 MATERIAL E MÉTODOS**

#### **3.1 Descrição do experimento**

O experimento foi realizado na área experimental pertencente ao setor de olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), município de Viçosa, Zona da Mata de Minas Gerais, coordenadas geográficas: 20° 45' de latitude sul e 42° 51' de longitude oeste, altitude média de 650 m. O experimento aconteceu no período de março a novembro de 2010.

Foram avaliados 30 acessos de alho, registrados no Banco de Germoplasma de Hortaliças (BGH/UFV). O delineamento experimental utilizado foi o de blocos completos casualizados, com quatro repetições.

As unidades experimentais foram constituídas por quatro fileiras transversais de um metro de comprimento, com plantas no espaçamento de 0,25 x 0,10 m, num total de 40 plantas, sendo consideradas como úteis as plantas das duas fileiras centrais. Depois de colhidas, as plantas foram submetidas ao processo de cura de campo e de galpão, por 3 e 60 dias respectivamente, expostas ao sol e em galpão. Após a cura, as plantas foram submetidas à toaleta cortando a parte aérea a 1,0 cm acima dos bulbos, retirando-se as raízes.

A avaliação foi feita sobre a matéria seca total da planta (MSTP), expressa em gramas por planta, obtida pelo somatório da matéria seca da folha (MSF), do pseudocaule (MSP), do bulbo (MSB) e da raiz (MSR). As MSF, MSP, MSB e MSR foram obtidas por secagem em estufa com ventilação forçada a 65°C até massa constante. A avaliação de matéria seca das partes das plantas desses acessos foi realizada nos quatro seguintes períodos: o período inicial (60 dias após plantio); segundo período (90 dias após plantio); terceiro período (120 dias após plantio); e o período final (150 dias após plantio).

Na Tabela 2 são apresentados os 30 acessos utilizados neste trabalho, com seus respectivos nomes comuns e origem. A estatística descritiva dos dados é apresentada na Tabela 3.

**Tabela 2.** Relação dos 30 acessos de alho registrados no Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV), seguidos de sua variedade e origem.

Nº	Acesso	Nome comum/variedade	Origem
1	4483	Barbado de Rio Grande	Florestal – MG
2	4491	Branco de Dourados	–
3	4499	Montes Claros	–
4	4500	Sacaia de Guanhões	–
5	4505	Patos de Minas	–
6	4567	Cateto Roxo	Gouvêia – MG
7	4814	–	Santa Tereza – ES
8	4815	–	S. Leopoldina – ES
9	4816	–	Rio Pantojas – ES
10	4817	–	Jequitibá – MG
11	4818	–	Santa Tereza – ES
12	4820	–	Santa Tereza – ES
13	4822	–	S. Leopoldina – ES
14	4823	–	S. Leopoldina – ES
15	4825	Sapé	Rio Pantojas – ES
16	4826	–	São Luís – MA
17	4829	Sapé	São Sebastião – SP
18	4835	–	S. Maria de Jetibá – ES
19	4837	–	Vargem Alegre – MG
20	4843	–	Itarana – ES
21	4844	–	Afonso Cláudio – ES
22	4899	–	–
23	5366	Cultura de tecidos	Brasília – DF
24	5934	Amarante	Viçosa – MG
25	5935	Caturra	Viçosa – MG
26	5938	Cateto Roxo	Viçosa – MG
27	5940	Amarante Novo Cruzeiro	Viçosa – MG
28	5946	Cateto Roxão	Viçosa – MG
29	5950	Chinês – 1	–
30	5951	Chinês – 3	Viçosa – MG



**Tabela 3.** Média, desvio padrão e valores mínimos e máximos da matéria seca total acumulada (em g) pelos 30 acessos de alho medida nos quatro períodos de tempo.

Dias após plantio	Média	Desvio padrão	Mínimo	Máximo
60	0,9730	0,3000	0,4464	1,6037
90	4,4448	0,9570	2,8650	6,7250
120	17,6591	3,5927	11,8000	28,7500
150	22,5479	4,6964	13,0250	36,1250

### 3.2 Modelos de regressão não linear

O modelo de regressão não linear utilizado foi o Logístico. Foram ajustados modelos para cada um dos acessos avaliados, utilizando todas as observações do acesso, e um modelo para todos os acessos utilizando as médias de cada acesso. O método de estimação para os parâmetros dos modelos de regressão não linear foi o de Mínimos Quadrados Ordinários, utilizando o método iterativo de Gauss-Newton. O modelo Logístico foi escolhido por apresentar melhores resultados para descrição de acúmulo de matéria seca e curvas de crescimento de plantas em diversos trabalhos (PÔRTO *et al.*, 2007; MAIA *et al.*, 2009; PUIATTI *et al.*, 2013; REIS *et al.*, 2014). O modelo Logístico é definido como:

$$y_i = \frac{\beta_1}{1 + \beta_2 e^{(-\beta_3 x_i)}} + \varepsilon_i ,$$

em que:

- $y_i$  é a i-ésima observação da variável resposta, a matéria seca total da planta (MSTP) expressa em gramas;
- $x_i$  é a i-ésima observação da variável preditora, representada pelos períodos da avaliação de matéria seca da planta (dias após plantio, DAP);
- $\beta_1$  é o parâmetro que representa o peso assintótico do acesso;

- $\beta_2$  é um parâmetro de locação que não possui interpretação biológica;
- $\beta_3$  é o parâmetro da taxa de crescimento dos acessos;
- $\varepsilon_i$  é o resíduo, que segue a distribuição assintótica  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ .

Um modelo de regressão quantílica não linear (R. Q.) foi ajustado de acordo com os quantis  $\tau = 0,25$ ,  $\tau = 0,5$  e  $\tau = 0,75$ , considerando todos os acessos. Os modelos foram ajustados pelo algoritmo de pontos interiores proposto por Koenker e Park (1996), que tem a finalidade de computar estimativas de regressão quantílica para casos em que a função resposta é não linear nos parâmetros. Foi também ajustado um modelo para cada acesso considerando apenas a mediana ( $\tau = 0,5$ ). O modelo é descrito por:

$$y_i = \frac{\beta_1(\tau)}{1 + \beta_2(\tau)e^{(-\beta_3(\tau)x_i)}} + \varepsilon_i(\tau) ,$$

em que:

- $\tau$  refere-se ao quantil assumido ( $\tau \in [0,1]$ );
- $\beta_1(\tau)$  é o parâmetro que representa o peso assintótico do acesso no quantil  $\tau$ ;
- $\beta_2(\tau)$  é um parâmetro de locação acesso no quantil  $\tau$ ;
- $\beta_3(\tau)$  é o parâmetro da taxa de crescimento dos acessos acesso no quantil  $\tau$ ;
- $\varepsilon_i(\tau)$  é o erro ou resíduo, supondo que os erros são independentes e identicamente distribuídos. Estes erros podem apresentar heterocedasticidade, uma vez que os parâmetros variam em função de  $\tau$ , mas segundo Koenker (2005), os modelos de regressão quantílica (R. Q.) podem incorporar a heterocedasticidade.

Para este trabalho, as estimativas iniciais dos parâmetros foram obtidas através da análise dos dados e das propriedades do modelo. Como  $\beta_1$  é a assíntota superior, uma aproximação viável para seu valor inicial é o maior valor observado da variável resposta.

### 3.2.1 Concordância entre modelos

Uma análise de regressão linear simples pode ser usada para verificar a existência de concordância entre dois modelos (MIRANDA *et al.*, 2014). Desta forma é possível testar se existe diferença significativa entre eles.

As variáveis do modelo de regressão linear simples,  $X_i$  e  $Y_i$ , são os valores preditos pelos modelos de regressão quantílica a serem comparados. Um modelo de regressão linear simples sem o intercepto é ajustado. Para verificar a concordância entre modelos é realizado o teste  $t$  para as hipóteses:

$$H_0 : \beta_1 = 1 \text{ vs } H_1 : \beta_1 \neq 1.$$

A estatística utilizada para testar estas hipóteses é:

$$t_{calc} = \frac{\hat{\beta}_1 - 1}{\sqrt{\frac{QMRes}{\sum_{i=1}^n X_i^2}}} \sim t_{(n-1)},$$

em que:

- $\hat{\beta}_1$  é a estimativa do coeficiente angular da regressão linear simples;
- $QMRes$  é o quadrado médio do resíduo da análise de variância da regressão linear simples;
- $n$  é o número de observações;
- e  $t_{calc}$  segue distribuição  $t$  de Student com  $n-1$  graus de liberdade.

Se o resultado do teste for não significativo, os modelos de regressão quantílica dos diferentes quantis são concordantes, ou seja, predizem a mesma resposta. Se a hipótese  $H_0$  for rejeitada, espera-se que a resposta predita pelos dois modelos de regressão quantílica seja diferente (existe diferença entre os quantis).

Neste trabalho, foram testadas as diferenças entre os modelos de regressão quantílica R. Q. ( $\tau=0,25$ ) e R. Q. ( $\tau=0,5$ ), e entre os modelos R. Q. ( $\tau=0,5$ ) e R. Q. ( $\tau=0,75$ ).

### 3.2.2 Avaliadores de qualidade de ajuste

Para comparar a qualidade de ajuste dos modelos foram calculados os seguintes avaliadores: quadrado médio do resíduo ( $QMR$ ); desvio médio absoluto dos resíduos ( $DMA$ ); critério de informação de Akaike ( $AIC$ ); e o coeficiente de determinação ( $R^2$  e  $R^1(\tau)$ ).

#### 3.2.2.1 Quadrado médio do resíduo

O quadrado médio do resíduo ( $QMR$ ) é calculado como:

$$QMR = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n - p},$$

em que:

- $y_i$  é o valor observado da matéria seca total da observação  $i$ ;
- $\hat{y}_i$  é o valor estimado da matéria seca total da observação  $i$  para dado modelo;
- $n$  é o número de observações;
- $p$  é o número de parâmetros do modelo.

Menor valor do quadrado médio do resíduo indica melhor ajuste do modelo.

### 3.2.2.2 Desvio médio absoluto dos resíduos

O desvio médio absoluto dos resíduos (*DMA*) é calculado como o somatório dos desvios entre os valores observados e estimados, dividido pelo número de observações (SARMENTO *et al.*, 2006):

$$DMA = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{n} ,$$

em que:

- $y_i$  é o valor observado da matéria seca total da observação  $i$ ;
- $\hat{y}_i$  é o valor estimado da matéria seca total da observação  $i$  para dado modelo;
- $n$  é o número de observações.

Quanto menor for o desvio médio absoluto dos resíduos, melhor será o ajuste do modelo aos dados.

### 3.2.2.3 Critério de informação de Akaike

Proposto por Akaike (1974), o critério de informação de Akaike (*AIC*), é uma estatística para comparação da qualidade de ajuste do modelo baseada no máximo da função de verossimilhança, e depende do número de observações e parâmetros do modelo a ser avaliado. Este critério admite a existência de um modelo real porém desconhecido, que descreve os dados. Ele tenta então escolher dentre um grupo de modelos aquele que minimiza a divergência de Kull-Leibler, medida relacionada à informação perdida por se usar o modelo aproximado ao invés do modelo “real”. O valor do *AIC* é calculado por:

$$AIC = -2\log L(\hat{\theta}) + 2k ,$$

em que:

- $k$  é o número de parâmetros no modelo;
- $L(\hat{\theta})$  é o máximo da função de verossimilhança para o modelo.

O melhor modelo segundo os critério de informação de Akaike é aquele que apresenta menor valor de  $AIC$ .

#### 3.2.2.4 Coeficiente de determinação

O coeficiente de determinação ( $R^2$ ) foi calculado como:

$$R^2 = 1 - \frac{SQR}{SQT} ,$$

em que:

- $SQR$  representa a soma de quadrados dos resíduos, calculada com base na diferença entre os valores estimados e observados da matéria seca total da planta de alho;
- $SQT$  representa a soma de quadrados total, com correção pela média, calculada como a diferença entre o valor observado e a média das observações da matéria seca total.

O coeficiente de determinação varia de 0 a 1, sendo que valores mais próximos de 1 representam melhor ajuste.

#### 3.2.2.5 Coeficiente de determinação para regressão quantílica

Koenker e Machado (1999) sugerem um critério análogo ao  $R^2$ , considerando os parâmetros estimados por um modelo de regressão quantílica. Enquanto o  $R^2$  mede o ajuste do modelo em relação à função condicional da média em termos da variância residual, este critério mede a qualidade de ajuste do modelo em um quantil específico em termos do somatório absoluto dos resíduos. Assim, esta medida avalia o ajuste do modelo apenas para dado quantil, e é calculada por:

$$R^1(\tau) = 1 - \frac{\hat{V}(\tau)}{\tilde{V}(\tau)},$$

em que:

- $\tau$  é o quantil do modelo ajustado;
- $\hat{V}(\tau)$  é função da variância dos resíduos do modelo ajustado para o quantil  $\tau$ ;
- $\tilde{V}(\tau)$  é a variância dos valores observados.

O valor deste critério está entre 0 e 1. Quanto maior o coeficiente de determinação  $R^1(\tau)$ , melhor a qualidade do modelo ajustado.

### 3.3 Distância entre acessos e modelos ajustados

Após ajustados os modelos de regressão quantílica para os 30 acessos, deseja-se classificar os acessos de acordo com os diferentes padrões de crescimento: acessos com diferentes pesos e taxas de crescimento possuem uma exigência nutricional variada, o que torna mais vantajoso distinguir seus tratamentos (SOUZA e MACÊDO, 2009). Foram calculadas as distâncias entre os valores observados de cada acesso e os valores preditos por cada quantil ( $\tau=0,25$ ,  $\tau=0,5$  e  $\tau=0,75$ ) dos modelos de regressão quantílica ajustados para os 30 acessos. A medida de dissimilaridade utilizada foi a distância euclidiana, calculada como:

$$d_{rs} = \sqrt{\sum_{j=1}^p (\hat{y}_{rj} - y_{sj})^2},$$

em que:

- $d_{rs}$  é a distância entre os valores preditos pelo modelo  $r$  e observados no acesso  $s$ ;

- $\hat{y}_{rj}$  é o valor predito da matéria seca da planta (em gramas) pelo modelo  $r$ ;
- $y_{sj}$  representa o valor observado da matéria seca da planta (em gramas) do acesso  $s$ ;
- $p$  é a dimensão da amostra (número de observações do acesso).

O acesso foi classificado conforme sua taxa de crescimento e peso assintótico, de acordo com a curva que apresenta menor distância entre valores preditos e observados: acessos de baixo interesse para o plantio ( $\tau = 0,25$ ); acessos de interesse intermediário ( $\tau = 0,5$ ); e acessos de alto interesse para o plantio ( $\tau = 0,75$ ).

### 3.4 Recursos computacionais

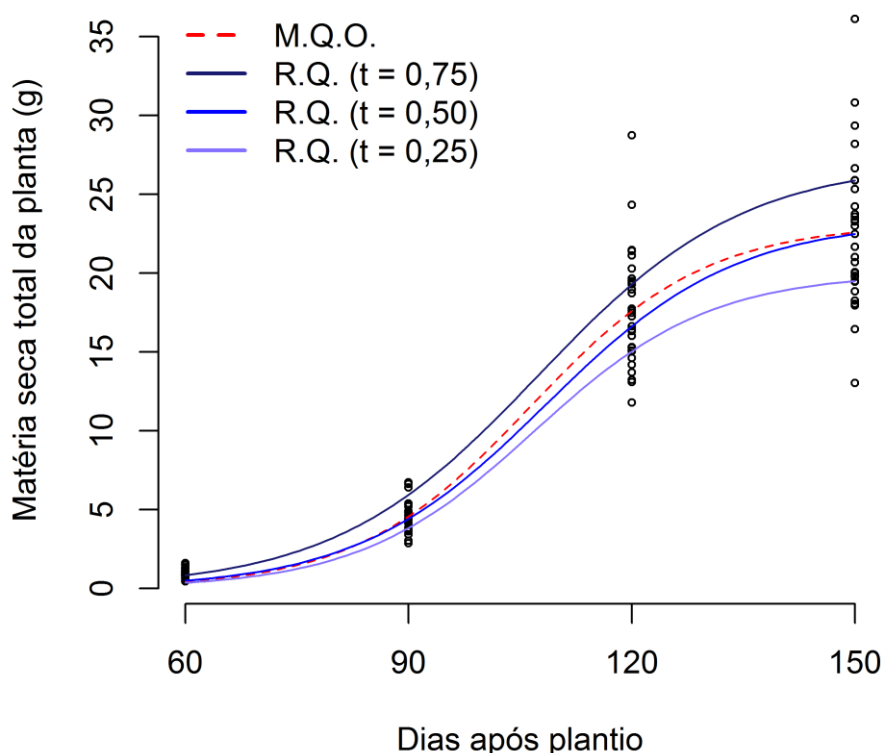
Todas as análises estatísticas deste trabalho foram realizadas utilizando-se o *software* R versão 3.2.1 (R DEVELOPMENT CORE TEAM, 2018). Para o ajuste dos modelos de regressão quantílica não linear foi utilizado o pacote *quantreg* (KOENKER, 2016). A função que realiza o ajuste de modelos de regressão não linear pelo método dos mínimos quadrados é parte do pacote padrão do *software* R.



## 4 RESULTADOS E DISCUSSÃO

### 4.1 Ajuste dos modelos de regressão não linear para todos os acessos

Considerando as médias das observações de cada acesso, foram ajustados quatro modelos não lineares. Primeiro foram ajustados três modelos de regressão quantílica, estimando seus parâmetros nos quantis  $\tau = 0,25$ ,  $\tau = 0,5$  e  $\tau = 0,75$ . O outro modelo foi ajustado pelo método dos Mínimos Quadrados Ordinários. As curvas dos modelos são apresentadas na Figura 1. Todas as curvas apresentaram um formato sigmoidal bem definido, característico em curvas de crescimento de plantas (PÔRTO *et al.*, 2007; MAIA *et al.*, 2009; REIS *et al.*, 2014).



**Figura 1.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para os 30 acessos.

Na Tabela 4 são apresentadas as estimativas dos parâmetros para os modelos ajustados, bem como a média, desvio padrão e coeficiente de variação. Os testes *t* para concordância entre modelos permitem concluir que

os três modelos de regressão quantílica diferem significativamente entre si ( $P < 0,001$ ), indicando que os acessos têm diferente impacto no acúmulo de matéria seca da planta.

**Tabela 4** Estimativas  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  e  $\hat{\beta}_3$ , seguidas pelo desvio padrão (DP), média, desvio e coeficiente de variação.

Modelos	Estimativas dos parâmetros		
	$\hat{\beta}_1$ (DP)	$\hat{\beta}_2$ (DP)	$\hat{\beta}_3$ (DP)
M. Q. O.	23,1320 (0,6991)	8694,29 (6733,11)	0,0852 (0,0076)
R. Q. ( $\tau = 0,25$ )	20,0051 (0,7189)	8739,27 (5606,66)	0,0849 (0,0072)
R. Q. ( $\tau = 0,5$ )	23,3309 (1,4246)	5077,55 (3209,59)	0,0787 (0,0074)
R. Q. ( $\tau = 0,75$ )	27,0572 (1,7074)	2434,28 (899,06)	0,0725 (0,0047)
Média	23,3813	6236,3475	0,0803
Desvio	2,8853	3060,7532	0,0060
CV (%)	12,3402	49,0792	7,4719

As estimativas  $\hat{\beta}_1$ , que representam o peso assintótico, foram próximas para os modelos de mínimos quadrados ordinários (M. Q. O.) e o de regressão quantílica na mediana (R. Q. ( $\tau = 0,5$ )), com massas de respectivamente 23,1320 g e 23,3309 g. O parâmetro  $\beta_1$  estimado no quantil  $\tau = 0,75$  apresentou o maior valor, representando os acessos com maior acúmulo de matéria seca (27,0572 g). O valor estimado para o quantil da R. Q. ( $\tau = 0,25$ ) possui menor peso assintótico (20,0051 g). Acessos que atingem maior massa seca no estágio fenológico final são mais promissores para o plantio, sendo estes os mais próximos da curva do quantil  $\tau = 0,75$ .

As estimativas do parâmetro  $\beta_2$  apresentaram a maior variação, mas como este não possui interpretação biológica direta, esses valores não representam problemas ou inconsistência nos modelos ajustados. Quanto às estimativas de  $\hat{\beta}_3$ , o quantil  $\tau = 0,25$  apresentou maior estimativa (0,0849), indicando que os acessos de menor peso final atingem a maturidade mais rapidamente, enquanto os modelos de produção mediana e maior ( $\tau = 0,5$  e

$\tau = 0,75$ , respectivamente) levam mais tempo para atingir o valor assintótico (0,0787 e 0,0725, respectivamente). A estimativa deste parâmetro para o modelo de M. Q. O. (0,0852) foi mais próxima da obtida pelo modelo R. Q. ( $\tau = 0,25$ ), sendo este o parâmetro que apresentou menor variação dentre os três estimados.

São apresentados na Tabela 5 os avaliadores de qualidade de ajuste calculados para cada modelo. Os valores em geral foram próximos para os diferentes modelos, com o quadrado médio do resíduo possuindo maior variação, enquanto o critério de Akaike foi o que menos variou.

**Tabela 5** Avaliadores de qualidade de ajuste: coeficiente de determinação ( $R^1(\tau)$  para modelos de regressão quantílica e  $R^2$  para mínimos quadrados), quadrado médio do resíduo ( $QMR$ ), desvio médio absoluto dos resíduos ( $DMA$ ) e critério de Akaike ( $AIC$ ) para os quatro modelos, seguidos de médias, desvio padrão e coeficiente de variação.

Modelos	Avaliadores			
	$R^1(\tau)$ ou $R^2$	$QMR$	$DMA$	$AIC$
M. Q. O.	0,9013	8,9957	1,9014	609,1170
R. Q. ( $\tau = 0,25$ )	0,8546	13,2490	2,1746	655,5780
R. Q. ( $\tau = 0,5$ )	0,8985	9,2476	1,8822	612,4302
R. Q. ( $\tau = 0,75$ )	0,8585	11,7785	2,3175	653,4408
Média	0,8782	10,8177	2,0689	632,6415
Desvio	0,0217	1,7762	0,1843	21,9123
CV (%)	2,4709	16,4193	8,9081	0,0346

Comparar a qualidade de ajuste entre modelos de diferentes quantis para os mesmos dados não faz sentido, mas a comparação entre o modelo ajustado pelo M. Q. O. e o modelo R. Q. ( $\tau = 0,5$ ) é possível, uma vez que o modelo de regressão quantílica para  $\tau = 0,5$  pode ser interpretado como uma regressão na mediana, enquanto o modelo de M. Q. O. utiliza a média, duas medidas de tendência central para representar os dados como um todo. Dentre

estes dois, o modelo de M. Q. O. apresenta melhores resultados para o coeficiente de determinação, quadrado médio do resíduo e critério de Akaike, e o modelo R. Q. ( $\tau = 0,5$ ) obteve melhores valores para desvio médio absoluto dos resíduos.

#### **4.2 Classificação dos acessos de acordo com sua taxa de crescimento e peso assintótico**

Os 30 acessos foram classificados de acordo com quantil do modelo de R. Q. que apresentou menor distância entre valor predito e observado. As classificações foram: baixo interesse para o plantio ( $\tau = 0,25$ ); médio interesse para o plantio ( $\tau = 0,5$ ); e alto interesse para o plantio ( $\tau = 0,75$ ). Em termos práticos, a matéria seca total acumulada reflete o potencial produtivo da planta, e também é um fator qualitativo por produzir bulbos de maior calibre, mais valorizados no mercado (DIRIBA-SHIFERAW, 2016). Portanto, acessos que apresentam maior peso ao final da colheita são mais promissores economicamente.

Na Tabela 6 são apresentadas as classificações de cada acesso, assim como a distância euclidiana entre os valores observados de cada acesso e preditos pelos três quantis do modelo de regressão quantílica. Dos 30 acessos, 12 apresentaram valores mais próximos dos estimados pelo modelo de quantil  $\tau = 0,25$ ; 6 foram mais próximos do modelo de quantil  $\tau = 0,5$ ; e os 12 restantes foram mais próximos do modelo de quantil  $\tau = 0,75$ .

Os acessos classificados como A (baixo interesse para o plantio) apresentam menor peso que os demais aos 150 dias após plantio (DAP), variando de 13,0250 g a 20,7000 g, enquanto apresentam maior taxa de crescimento. Neste grupo estão o acesso de menor peso (acesso 22) e o acesso que atinge o peso máximo mais rapidamente (acesso 27). Os acessos de classificação B (interesse intermediário) apresentam peso aos 150 DAP em geral entre as duas classificações, variando de 20,0000 g a 23,7500 g. Sua taxa de crescimento é menor que dos de classificação A e pouco menor que dos de classificação C.

**Tabela 6** Classificação dos 30 acessos de alho e distância euclidiana entre os valores observados em cada acesso e preditos pelos três modelos de regressão quantílica.

Acesso	Quantil			Classificação
	$\tau = 0,25$	$\tau = 0,5$	$\tau = 0,75$	
1	0,1948	3,4757	8,0140	A
7	1,9261	4,5881	8,7603	A
12	0,9330	3,9746	8,4087	A
13	0,9832	3,4485	7,8404	A
15	1,5968	2,4477	6,7124	A
19	3,3912	5,7568	10,1586	A
21	2,3492	5,6329	10,1570	A
22	6,7786	10,1103	14,4170	A
24	1,4718	4,6810	9,0747	A
25	1,4514	4,0328	8,5464	A
27	3,3446	6,6795	11,1190	A
28	2,0965	2,4819	6,5687	A
<hr/>				
2	4,5794	1,5676	4,1193	B
8	2,7858	2,7378	6,3400	B
10	4,2713	0,9383	3,6931	B
20	2,0914	1,8220	6,2217	B
29	4,3203	2,2406	4,5070	B
30	4,9340	1,7828	3,1753	B
<hr/>				
3	7,3814	4,0787	1,7136	C
4	11,8375	8,4373	4,1483	C
5	21,7391	18,3983	13,9788	C
6	11,8081	8,8136	5,2493	C
9	10,0389	6,6763	2,5681	C
11	6,1041	3,1517	2,8669	C
14	8,2210	5,3319	2,5678	C
16	11,6559	8,4215	5,3862	C
17	7,7016	4,3567	1,7319	C
18	6,0881	3,2317	2,9424	C
23	7,1110	4,9067	4,5826	C
26	8,3523	5,0403	1,4490	C

A: baixo interesse para plantio; B: interesse intermediário plantio; C alto interesse para plantio.

Já os acessos de classificação C (alto interesse para o plantio) foram caracterizados por maiores valores de matéria seca total da planta aos 150 DAP e menores valores para taxa de crescimento. O peso dos acessos com esta classificação está entre 22,4750 g e 36,1250 g. O acesso com maior peso final está neste grupo (Acesso 5).

Os grupos formados são similares com os do trabalho de REIS *et al.* (2014), que ajustaram modelos não lineares para descrever o acúmulo de matéria seca em grupos de acessos de alho providos do mesmo experimento. A regressão quantílica apresentou vantagens para ajuste de modelos não lineares para curvas de acúmulo de matéria seca: enquanto no trabalho de REIS *et al.* (2014) foi necessária uma análise de agrupamento para obter grupos de acessos mais homogêneos antes de se ajustar os modelos para cada grupo, a metodologia da regressão quantílica permite ajustar modelos em quantos quantis for de interesse, tornando a etapa de classificação mais simples. A regressão quantílica utilizou todos os dados para o ajuste dos modelos, permitindo uma análise mais completa da amostra.

#### **4.3 Ajuste dos modelos de regressão não linear para cada acesso**

Para cada um dos acessos em estudo, foi ajustado um modelo pelo método dos mínimos quadrados. As estimativas dos parâmetros dos modelos são apresentadas na Tabela 7. A variedade entre os acessos é notável ao observar as estimativas  $\hat{\beta}_1$ . As estimativas dos pesos assintóticos variam de 13,1714 g a 36,8230 g. As estimativas  $\hat{\beta}_3$  apresentam coeficiente de variação um pouco maior, com valores indo de 0,0569 a 0,1647. As estimativas do parâmetro  $\beta_2$  foram as que apresentaram maior variação.

**Tabela 7** Estimativas  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  e  $\hat{\beta}_3$  para os modelos ajustados pelo método dos mínimos quadrados para os 30 acessos.

Estimativas dos parâmetros				Estimativas dos parâmetros			
Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	19,9138	11136,1306	0,0871	16	35,7194	1809,8259	0,0623
2	24,8225	7882,5965	0,0806	17	26,6571	11791,0370	0,0861
3	25,9547	13738,1344	0,0885	18	23,5964	25242,0750	0,0976
4	30,5295	6141,3567	0,0797	19	21,9687	3003,3144	0,0680
5	36,8230	18714,2547	0,0924	20	21,4640	21150,0729	0,0925
6	26,7712	358490,2244	0,1255	21	18,8893	3446,8393	0,0749
7	18,1089	46979,1433	0,1063	22	13,1714	6588488,9485	0,1647
8	20,1383	77884,8385	0,1103	23	22,5216	4462035,1298	0,1525
9	31,6261	616,8951	0,0569	24	18,7351	4604,4808	0,0808
10	23,8497	4111,0721	0,0774	25	21,1029	2777,7372	0,0712
11	23,9633	23223,5409	0,0956	26	26,4732	12289,1350	0,0883
12	19,2031	14319,9648	0,0909	27	16,6024	62875,0920	0,1066
13	20,2784	3132,6655	0,0756	28	22,0721	1001,7167	0,0643
14	24,5586	9966,1049	0,0915	29	21,8656	45265,5515	0,1041
15	20,4132	13836,8669	0,0912	30	24,6123	1848,4903	0,0700

Dos acessos analisados por estes modelos, destacam-se o acesso 5 (BGH 4505), que apresenta maior peso assintótico e uma taxa de crescimento pouco acima da média, e o acesso 22 (BGH 4899), que apesar de possuir menor peso assintótico é o que atinge a maturidade mais rapidamente. O acesso 22 possui alguns valores extremos em suas observações, o que é refletido nas estimativas dos parâmetros do modelo ajustado para este. Muitos autores optam por remover valores extremos para evitar inconvenientes provocados por estes, mas isso significa sacrificar informações no conjunto de dados, uma vez que o *outlier* faz parte do fenômeno em estudo (LY *et al.*, 2013; BARROSO *et al.*, 2015).

Além da existência de possíveis valores extremos, ainda existe a questão da heterocedasticidade: as variâncias não são iguais para todas as observações. Conforme a planta cresce as observações visivelmente se distanciam, resultando em uma variância nos últimos períodos maior que nos

períodos iniciais. Isso não representa problema para o caso de regressão não linear quando se tem uma amostra grande o suficiente, pois a inferência neste caso fundamenta-se na teoria assintótica (GUJARATI, 2011). Alguns acessos também apresentam observações assimétricas. O uso de metodologias tradicionais quando os dados apresentam assimetria e valores discrepantes podem levar a resultados incorretos (BARROSO *et al.*, 2015). Assim, a regressão quantílica pode ser mais robusta estimando a mediana, ao invés da média utilizada nos mínimos quadrados ordinários.

Na Tabela 8 são apresentadas as estimativas dos parâmetros para os modelos de regressão quantílica ( $\tau=0,5$ ) ajustados para cada um dos acessos. A média da estimativa do parâmetro  $\beta_1$  foi 24,4780, com valores de peso assintótico variando de 15,2428 g a 39,2031 g. Este parâmetro foi também o que apresentou menor variação dentre os três. As estimativas da taxa de crescimento  $\hat{\beta}_3$  foram de 0,0517 a 0,1233, com média de 0,0789. Estimativas do parâmetro  $\beta_2$  novamente apresentaram maior coeficiente de variação que os demais. Se comparados com os parâmetros dos modelos de M. Q. O., os parâmetros estimados pelos modelos R. Q. ( $\tau=0,5$ ) apresentaram menores coeficientes de variação para  $\hat{\beta}_1$  e  $\hat{\beta}_2$ , mas maior variação para  $\hat{\beta}_3$ .

Para estes modelos, o acesso 5 (BGH 4505) também foi o que apresentou maior peso assintótico, mas possui uma das menores taxas de crescimento se comparado com as demais estimativas. Ele é seguido pelos acessos 9 (BGH 4816) e 6 (BGH 4567), com respectivamente 33,6592 g e 31,1177 g, sendo que o acesso 9 apresenta taxa de crescimento próxima à do acesso 5, enquanto o acesso 6 apresenta uma das maiores taxas. O acesso 27 (BGH 5940) possui a maior taxa de crescimento. O acesso 22 (BGH 4899) apresenta o menor peso assintótico.

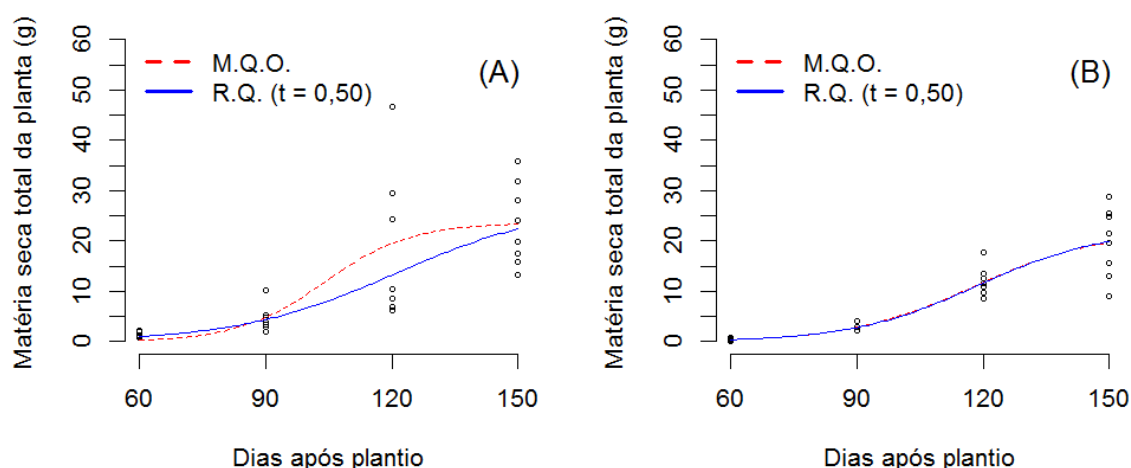


**Tabela 8** Estimativas  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  e  $\hat{\beta}_3$  para os modelos de regressão quantílica ( $\tau = 0,5$ ) ajustados para os 30 acessos.

Estimativas dos parâmetros				Estimativas dos parâmetros			
Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	20,8701	5112,6058	0,0778	16	30,7094	2461,1314	0,0672
2	29,8621	2187,9770	0,0644	17	25,3029	60787,1650	0,1029
3	29,7451	2122,6534	0,0666	18	27,2990	608,0382	0,0529
4	30,0390	1366,6752	0,0636	19	22,4065	2760,8232	0,0667
5	39,2031	667,9776	0,0541	20	25,9023	751,5820	0,0541
6	31,1177	257169,0165	0,1202	21	21,3264	1352,0442	0,0642
7	18,3667	105149,3597	0,1145	22	15,2428	5912,2971	0,0854
8	19,9983	2319,8376	0,0720	23	24,0434	119195,8615	0,1120
9	33,6592	464,9228	0,0517	24	20,7961	1925,5455	0,0680
10	26,2894	1671,9705	0,0640	25	22,4941	955,2751	0,0600
11	25,7614	1789,5481	0,0649	26	27,0578	30470,0656	0,0998
12	17,4084	32196,4634	0,1030	27	16,3483	319238,2345	0,1233
13	21,5734	18490,5154	0,0945	28	22,8523	757,7538	0,0600
14	24,2456	2262,9481	0,0742	29	21,5339	27954,0348	0,0983
15	19,3306	45248,7818	0,1047	30	23,5553	945,6884	0,0633

Comparando os modelos obtidos pelos dois métodos, as estimativas do peso assintótico variaram menos. Algumas estimativas foram bem próximas, sendo que o modelo R. Q. ( $\tau = 0,5$ ) apresentou maiores valores para este parâmetro na maioria dos acessos, enquanto as estimativas de M. Q. O. foram menores, possivelmente subestimando o peso final dos acessos. As estimativas dos parâmetros  $\beta_2$  variaram consideravelmente. Para maioria dos modelos, as estimativas de M. Q. O. para este parâmetro foram bem maiores, chegando a diferenças muito grandes em alguns acessos (acesso 22, por exemplo), mas este é apenas um parâmetro de locação e isso não interfere na interpretação dos demais. Em relação à estimativa da taxa de crescimento, os modelos de M. Q. O. apresentaram maiores valores para a maioria dos acessos, afirmando que o amadurecimento ocorre mais cedo que nas estimativas de R. Q. ( $\tau = 0,5$ ).

A Figura 2 mostra a diferença dos modelos de M. Q. O. e R. Q. ajustados para os mesmos acessos. O acesso 18 (BGH 4835) possui grande assimetria nos dados, principalmente aos 120 DAP, incluindo um valor discrepante neste período. Algumas observações distantes influenciam fortemente a média, e consequentemente a estimação dos parâmetros. A mediana é uma medida de tendência central mais robusta, sendo mais apropriada quando os dados apresentam tamanhas variações (KOENKER e BASSET, 1978). No caso do acesso 19 (BGH 4837), os dados são mais simétricos e homogêneos; como resultado, as curvas ajustadas para este acesso por ambas as abordagens são muito próximas.



**Figura 2.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para os acessos de alho 18 (A) e 19 (B).

Na Tabela 9 são apresentados os avaliadores de qualidade de ajuste para os modelos ajustados pelo método dos mínimos quadrados. O critério de Akaike foi o avaliador com menor variação para estes modelos, com  $162,9141 < AIC < 253,8458$ . A variação do coeficiente de determinação também foi pequena, com  $55,65\% < R^2 < 88,43\%$ . O quadrado médio do resíduo foi o avaliador com maior variação, com valores entre 8,1795 e 140,2244. O desvio médio absoluto dos resíduos teve uma variação intermediária, com  $1,7608 < DMA < 7,4160$ .

**Tabela 9** Avaliadores de qualidade de ajuste: coeficiente de determinação ( $R^2$ ), quadrado médio do resíduo ( $QMR$ ), desvio médio absoluto dos resíduos ( $DMA$ ) e critério de Akaike ( $AIC$ ) para os modelos ajustados pelo método dos mínimos quadrados para os 30 acessos.

Acesso	Avaliadores			
	$R^2$	$QMR$	$DMA$	$AIC$
1	0,8059	16,3468	2,5101	185,0710
2	0,7816	26,8889	3,2506	200,9968
3	0,7222	43,5013	4,1120	216,3913
4	0,8230	31,6188	3,8449	206,1820
5	0,6246	140,2244	7,4160	253,8458
6	0,7892	37,6676	4,2579	211,7836
7	0,7604	18,8450	2,8089	189,6219
8	0,8176	17,1240	2,7236	186,5574
9	0,6838	55,4589	4,7923	224,1625
10	0,8217	19,6205	2,8864	190,9123
11	0,7661	30,0073	3,5221	204,5080
12	0,8843	8,1795	1,7608	162,9141
13	0,7803	17,4366	2,7833	187,1362
14	0,8358	20,2500	3,0773	191,9230
15	0,8231	15,4772	2,6696	183,3217
16	0,7867	41,4911	4,2717	214,8773
17	0,7858	31,8982	3,3192	206,4636
18	0,5952	66,2501	5,2111	229,8519
19	0,8301	13,1830	2,0001	178,1877
20	0,5565	63,3592	4,3900	228,4242
21	0,8041	13,2277	2,2518	178,2961
22	0,6994	14,0044	2,5022	180,1219
23	0,8780	14,9358	2,2018	182,1823
24	0,7733	16,6311	2,7178	185,6228
25	0,8547	10,9338	2,2289	172,2013
26	0,8787	16,6556	2,8101	185,6698
27	0,7379	17,6767	2,7026	187,5739
28	0,7597	20,9509	2,9971	193,0118
29	0,8704	13,2719	2,2801	178,4028
30	0,7990	21,9474	3,2344	194,4987

O modelo ajustado para o acesso 12 (BGH 4820) foi o melhor avaliado de acordo com todos os quatro critérios. O modelo ajustado para o acesso 20 (BGH 4843) recebeu pior classificação segundo o avaliador  $R^2$ , enquanto o modelo ajustado para o acesso 5 (BGH 4505) recebeu pior classificação para os demais avaliadores.

Na Tabela 10 são apresentados os avaliadores de qualidade de ajuste para os modelos de regressão não linear quantílica ( $\tau=0,5$ ) ajustados. O critério de informação de Akaike teve menor variação dentre estes modelos, com valores de 147,1344 a 238,1785. O coeficiente  $R^1(\tau=0,5)$  também apresenta coeficiente de variação relativamente baixo, com mínimo e máximo de respectivamente 43,71% e 87,33. O desvio médio absoluto variou de 1,6688 a 6,9216, e o quadrado médio do resíduo de 8,9521 a 167,1717, sendo este último o que apresentou maior variação.

O modelo ajustado para o acesso 12 (BGH 4820) foi novamente o melhor avaliado segundo todos os critérios. As observações deste acesso apresentaram em geral menor variação que os demais, principalmente aos 120 DAP, em que a planta já está próxima de atingir o peso final. O modelo que recebeu os piores valores dos avaliadores  $QMR$ ,  $DMA$  e  $AIC$  foi o ajustado para o acesso 5 (BGH 4505), apresentando também um valor baixo de  $R^1(\tau=0,5)$ . Isso aconteceu possivelmente pelo fato de este acesso apresentar os maiores valores de matéria seca, o que o penalizou mais pela maior distância entre as observações. Este acesso apresentou alta variação nas suas observações, se comparado com os demais. O modelo ajustado para o acesso 18 (BGH 4835) recebeu o menor valor do avaliador  $R^1(\tau=0,5)$ .

**Tabela 10** Avaliadores de qualidade de ajuste: coeficiente de determinação ( $R^1(\tau = 0,5)$ ), quadrado médio do resíduo ( $QMR$ ), desvio médio absoluto dos resíduos ( $DMA$ ) e critério de Akaike ( $AIC$ ) para os modelos de regressão quantílica ( $\tau = 0,5$ ) ajustados para os 30 acessos.

Acesso	Avaliadores			
	$R^1(\tau)$	$QMR$	$DMA$	$AIC$
1	0,8034	16,5568	2,5012	173,0336
2	0,7640	29,0589	3,1982	188,7674
3	0,7087	45,6194	3,9941	202,9898
4	0,7860	35,8457	3,5756	195,9064
5	0,5600	167,1717	6,9216	238,1785
6	0,7464	45,3043	4,0620	204,0687
7	0,7589	18,9582	2,7756	179,6966
8	0,7791	20,7400	2,5338	173,8633
9	0,6641	56,2794	4,7141	213,5966
10	0,8044	21,5332	2,7490	179,0813
11	0,7218	35,6920	3,3399	191,5422
12	0,8733	8,9521	1,6688	147,1344
13	0,7431	20,3905	2,6734	177,2959
14	0,8215	22,1718	2,9569	183,7459
15	0,8202	15,7304	2,6020	175,5640
16	0,7605	44,0084	4,1325	205,1699
17	0,7825	32,3926	3,1984	188,7712
18	0,4371	77,5398	5,0697	218,2513
19	0,8301	13,1971	1,9928	158,4925
20	0,5124	68,1786	4,2322	206,6954
21	0,8003	13,9329	2,1481	163,2955
22	0,6776	15,0181	2,3549	169,1770
23	0,8716	15,7209	2,1690	163,9156
24	0,7622	17,4479	2,6730	177,2858
25	0,8515	11,1135	2,1800	164,2382
26	0,8643	18,6252	2,6657	177,1118
27	0,7365	17,7740	2,6586	176,9401
28	0,7580	21,0954	2,9393	183,3637
29	0,8678	13,5363	2,2356	165,8501
30	0,7816	23,0755	3,1509	187,8142

Comparando os avaliadores calculados para os modelos ajustados pelos dois métodos para cada acesso, nota-se que os modelos de M. Q. O. apresentaram melhores valores segundo o coeficiente de determinação e quadrado médio do resíduo, enquanto os modelos de R. Q. foram mais bem avaliados de acordo com o desvio médio absoluto dos resíduos e o critério de Akaike (Tabela 9 e Tabela 10). Esse resultado é esperado, uma vez que o método dos M. Q. O. obtém os estimadores que minimizam a soma dos quadrados dos resíduos da regressão, e os avaliadores  $R^2$  e  $QMR$  se baseiam na soma do quadrado dos resíduos. O critério  $DMA$  considera a soma das distâncias entre valores estimados e observados, e seu valor mínimo é obtido na mediana ( $\tau = 0,5$ ).

## 5 CONCLUSÃO

Foi possível ajustar um modelo de regressão quantílica não linear Logístico que permite distinguir os acessos de acordo com os diferentes níveis do acúmulo de matéria seca ao longo do tempo. Os 30 acessos foram classificados de acordo com o quantil de estimativas mais próximas, dos quais 12 foram classificados como de menor interesse para plantio (menor valor de peso assintótico, mas maior taxa de crescimento); seis foram classificados como intermediários; e 12 foram classificados como de maior interesse para plantio, possuindo menor taxa de crescimento e maior peso assintótico.

A regressão quantílica mediana foi eficiente para ajustar modelos de regressão não linear para descrição do acúmulo de matéria seca ao longo do tempo, gerando estimativas de parâmetros mais uniformes e robustas mesmo quando os dados apresentam assimetria, variâncias heterogêneas ou valores discrepantes.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

AKAIKE, H. A. New Look at the Statistical Model Identification. **IEEE Transactions on automatic control**, Minato-ku, Tokyo, Japan. v.19, n.6, p.716- 723, Dec. 1974.

ANJO, D. F. C. Alimentos funcionais em angiologia e cirurgia vascular. **J. Vasc. Br.**, v. 3, n. 2, p. 145-154, 2004.

BARROSO, L. M. A.; NASCIMENTO, M.; NASCIMENTO, A. C. C., SILVA, F. F. e; CRUZ, C. D.; BHERING, L. L.; FERREIRA, R. de P. Metodologia para análise de adaptabilidade e estabilidade por meio de regressão quantílica. **Pesquisa Agropecuária Brasileira**, Brasília, v. 50, n. 4, p. 290-297. 2015.

BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: John Wiley and Sons, 1988.

BOUYÉ, E.; SALMON, M. Dynamic copula quantile regressions and tail area dynamic dependence in Forex markets. **The European Journal of Finance**, 15 (7-8), 721-750. 2009.

BROWN, J. E.; FITZHUGH JUNIOR, H. A.; CARTWRIGHT, T. C. A comparison of nonlinear models for describing weight-age relationships in cattle. **Journal of Animal Science**, Champaign, v. 42, n. 4, p. 810-818, Apr. 1976.

CADE, B. S.; NOON, B. R. A gentle introduction to quantile regression for ecologists, **Frontiers in Ecology and the Environment**, v. 1, n. 8, p. 412–420, 2003.

CAO, Q. V.; DEAN, T. J. Using nonlinear quantile regression to estimate the self-thinning boundary curve. 17th biennial southern silvicultural research conference. **Proceedings**. e–Gen. Tech. Rep. SRS–203. Asheville, NC: U.S.



Department of Agriculture, Forest Service, Southern Research Station. 3 p. 2015.

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

DENISE, R. S. K.; BRINKS, J. S. Genetic and environmental aspects of the growth curve parameters in beef cows. **Journal of Animal Science**, Champaign, v. 61, n. 6, p. 1431-1440, July 1985.

DIRIBA-SHIFERAW, G. Review of Management Strategies of Constraints in Garlic (*Allium sativum* L.) Production. **Journal of Agricultural Sciences**. 11(3), pp.186–207. 2016.

FERNANDES, T. J.; MUNIZ, J. A.; PEREIRA, A. A.; MUNIZ, F. R.; MUIANGA, C. A. Parameterization effects in nonlinear models to describe growth curves. **Acta Scientiarum**, Maringá, v. 37, n. 4, p. 397-402, 2015.

GUJARATI, D. N. **Econometria Básica**. 4. ed. São Paulo: Pearson Makron Books, 920p., 2011.

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

HUNTER, D. R.; LANGE, K. Quantile Regression via an MM Algorithm. **Journal of Computational and Graphical Statistics**, 9: 60–77. 1999.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Levantamento Sistemático da Produção Agrícola. Rio de Janeiro, v.29 n.02 p.1-79 fevereiro. 2016. Disponível em: <[ftp://ftp.ibge.gov.br/Producao\\_Agricola/Levantamento\\_Sistematico\\_da\\_Producao\\_Agricola\\_%5Bmensal%5D/Fasciculo/2016/lspa\\_201602.pdf](ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_%5Bmensal%5D/Fasciculo/2016/lspa_201602.pdf)>. Acesso em: 16/05/2017.

KOENKER R. quantreg: Quantile Regression. R package version 5.21, 2016. Disponível em: <<https://cran.r-project.org/web/packages/quantreg/index.html>>. . Acesso em: 17/04/2016.

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

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

KOENKER, R.; HALLOCK, K. F. Quantile Regression. **Journal of Economic Perspectives**, v. 15, n. 4, p. 143–156, Fall 2001.

KOENKER, R.; MACHADO, J. A. F. Goodness of Fit and Related Inference Processes for Quantile Regression. **Journal of the American Statistical Association**, v. 94, n. 448, p.1296-1310, 1999.

KOENKER, R.; PARK, B. J. An interior point algorithm for nonlinear quantile regression. **Journal of Econometrics**, 71, 265-283. 1996.

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 x Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava. **Crop Science**. v.53, p. 1312-1325, 2013.

MAIA, E.; SIQUEIRA, D. L.; SILVA, F. F.; PETERNELLI, L. A.; SALOMÃO, L. C. C. Método de comparação de modelos de regressão não-lineares em bananeiras. **Ciência Rural**, Santa Maria, v.39, n.5, p.1380-1386, 2009.

MARCHIORI, V. F. **Propriedades funcionais do alho (Allium Sativum L.)**. 2005. Disponível em: <[http://www.esalq.usp.br/siesalq/pm/alho\\_revisado.pdf](http://www.esalq.usp.br/siesalq/pm/alho_revisado.pdf)>. Acesso em: 26/8/2015.

MAZUCHELI, J.; ACHCAR, J. A. Algumas considerações em regressão não linear. **Acta Scientiarum**, Maringá, v.24, n.6, p.1761-1770, 2002.

MAZZINI, A. R. A.; MUNIZ, J. A.; SILVA, F. F. e; AQUIN, L. H. Curvas de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 422-427, 2005.

MEKETON, M. **Least absolute value regression**, AT&T Bell Laboratories Technical Report, Holmdel, NJ. 1986.

MIRANDA, A. C. R., PRUSKI, F., MARTINEZ, M. A., CECON, P. R. Métodos de Separação dos Escoamentos Superficial Direto e Subterrâneo: Estudo de Caso para a Bacia do Rio das Velhas. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v. 19, n. 2, p. 169-181, 2014.

MOSTELLER, F; TUKEY, J. W. **Data Analysis and Regression: A Second Course in Statistics**, Addison-Wesley. 1977.

MOTULSKY, H.; CHRISTOPOULOS, A. **Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting**. 4.ed. San Diego, CA:GraphPad Software, 351p., 2003.

MUELLER, S.; BIASI, J.; MENEZES SOBRINHO, J. A.; MULLER, J. J. V. Comportamento de cultivares de alho, plantio de junho. **Pesquisa Agropecuária Brasileira**, v.25, n.11, p.1561-1567. 1990.

MUGGEO, V. M. R.; SCIANDRA, M.; TOMASELLO, A; CALVO, S. Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. **Environmental and Ecological Statistics**, v. 20, n. 4, p. 519-531. 2013.

POLLICE, A.; MUGGEO, V.; TORRETTA, F.; BOCHICCHIO, R.; AMATO, M. Growth curves of sorghum roots via quantile regression with P-splines. In 47th

Scientific Meeting of the Italian Statistical Society. **Proceedings**. Cagliari: Cooperativa Universitaria Editrice Cagliariitana. 2014.

PÔRTO, D. R. de Q.; CECILIO FILHO, A. B.; MAY, A.; VARGAS, P. F. Acúmulo de macronutrientes pela cultivar de cebola Superex estabelecida por semeadura direta. **Ciência Rural**, v.37, p.949-955, 2007.

PRUDENTE, A. A. **Modelos não-lineares de regressão: alguns aspectos de teoria assintótica**. Dissertação (Mestrado em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco. Recife, 108f, 2009.

PUIATTI, G. A.; CECON, P. R.; NASCIMENTO, M.; PUIATTI, M.; FINGER, F. L.; SILVA, A. R. da. Análise de agrupamento em seleção de modelos de regressão não lineares para descrever o acúmulo de matéria seca em plantas de alho. **Rev. Bras. Biom.**, São Paulo, v.31, n3, p.337-351, jul.-set. 2013.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2017. Disponível em <<http://www.R-project.org>>. Acesso em 23/05/2017.

RATKOWSKY, D. A. **Nonlinear regression modeling: a unified practical approach**. Marcel Dekker, New York. 1983.

REGAZZI, A. J. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não linear. **Revista Ceres**, Viçosa, v. 50, n. 287, p. 9-26, 2003.

REIS, R. M.; CECON, P. R.; PUIATTI, M.; FINGER, F. L.; NASCIMENTO, M.; SILVA, F. F.; CARNEIRO, A. P. S.; SILVA, A. R. Modelos de regressão não linear aplicados a grupos de acessos de alho. **Horticultura Brasileira**, Vitória da Conquista, v. 32, n. 2, p. 178-183, 2014.

RESENDE, F. V.; FAQUIN, V.; SOUZA, R. J. de; SILVA, V. S. Acúmulo de matéria seca e exigências nutricionais de plantas de alho provenientes de cultura de tecidos e de propagação convencional. **Horticultura Brasileira**, Brasília, v. 17 n. 3, p. 220-226, 1999.

SARMENTO, J. L. R.; REGAZZI, A. J.; SOUZA, W. H.; TORRES, R. A.; BREDI, F.C.; MENEZES, G. R. O. Estudo da curva de crescimento de ovinos Santa Inês. **Revista Brasileira de Zootecnia**, v. 35, n. 2, p. 435-442, 2006.

SCHRÖDER, H. K.; ANDERSEN, H. E.; KIEHL, K. Rejecting the Mean: Estimating the Response of Fen Plant Species to Environmental Factors by Non-Linear Quantile Regression. **Journal of Vegetation Science**, v. 16, n. 4, p. 363-378, 2005.

SHARMA, B. D.; KAUL, H. N.; SINGH, M. Growth analysis of potato varieties in autumn in subtropical conditions. **New Botanist**, Lucknow, v. 20, n. 54, p. 55-64, 1993.

SILVA, E. N. da; PORTO JÚNIOR, S. da S. Sistema financeiro e crescimento econômico: uma aplicação de regressão quantílica. **Economia Aplicada**, Ribeirão Preto, v. 10, n. 3, p. 425-442, 2006.

SILVA, H. R. F. MELO, V. L.; PACHECO, D. D; ASSIS, Y. J. M. de; SALES, H. dos R. Acúmulo de matéria seca e micronutrientes em mandioca consorciada com bananeira. **Pesquisa Agropecuária Tropical**, Goiânia, v. 44, n. 1, p. 15-23. 2014.

SILVEIRA, F. G.; SILVA, F. F.; CARNEIRO, P. L. S.; MALHADO, C. H. M.; MUNIZ, J. A. Análise de agrupamento na seleção de modelos de regressão não-lineares para curvas de crescimento de ovinos cruzados. **Ciência Rural**, Santa Maria, v.41, n.4, p.692-698, 2011.

SORRELL, B. K; TANNER, C. C; BRIX, H. Regression analysis of growth responses to water depth in three wetland plant species. **AoB Plants**. 2012.

SOUSA, I. F.; NETO, J. E. K.; MUNIZ, J. A.; GUIMARÃES, R. M.; SAVIAN, T. V.; MUNIZ, F. R. Fitting nonlinear autoregressive models to describe coffee seed germination. **Ciência Rural**, Santa Maria, v. 44, n. 11, 2014.

SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear**. 1 ed. Brasília: Embrapa, 489p. 1998.

SOUZA, R. J. de; MACÊDO, F. S. **Cultura do alho: tecnologias modernas de produção**. Lavras: Editora UFLA, 181p. 2009.

TRANI, P. E. **Cultura do alho (*Allium sativum*): Diagnóstico e recomendações para seu cultivo no Estado de São Paulo**. 2009. Artigo em Hypertexto. Disponível em: <[http://www.infobibos.com/Artigos/2009\\_2/alho/index.htm](http://www.infobibos.com/Artigos/2009_2/alho/index.htm)>. Acesso em: 26/8/2017.

## **APÊNDICE**

## A – Análises de concordância entre modelos

Na Tabela 1A são apresentadas as estimativas de  $\hat{\beta}_1$  dos modelos lineares ajustados para as comparações entre os modelos R. Q. ( $\tau = 0,25$ ) e R. Q. ( $\tau = 0,5$ ), e entre os modelos R. Q. ( $\tau = 0,75$ ) e R. Q. ( $\tau = 0,5$ ).

**Tabela 1A** Resultados da comparação entre os três modelos de regressão quantílica.

Modelos	$\hat{\beta}_1$
R. Q. ( $\tau = 0,25$ ) x R. Q. ( $\tau = 0,5$ )	0,8803 *
R. Q. ( $\tau = 0,75$ ) x R. Q. ( $\tau = 0,5$ )	1,1592 *

(\*) significativo a 1% de probabilidade.

Na Tabela 2A são apresentados os resíduos das regressões para as comparações entre os modelos R. Q. ( $\tau = 0,25$ ) e R. Q. ( $\tau = 0,5$ ), e para comparação entre os modelos R. Q. ( $\tau = 0,75$ ) e R. Q. ( $\tau = 0,5$ ).

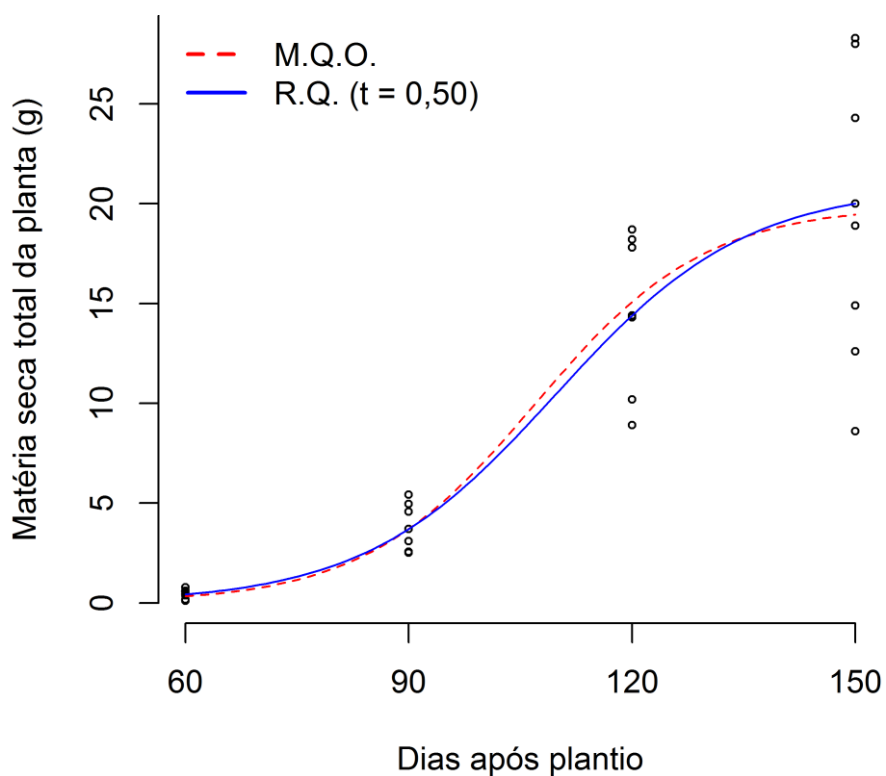
**Tabela 2A** Resíduos das regressões na comparação entre os três modelos de regressão quantílica.

Comparação: R. Q. ( $\tau = 0,25$ ) x R. Q. ( $\tau = 0,5$ ) R. Q. ( $\tau = 0,75$ ) x R. Q. ( $\tau = 0,5$ )				
F. V.	G. L.	Q. M.	G. L.	Q. M.
Regressão	1	18629,5145	1	32304,0050
Resíduo	118	0,0637	118	0,1920
Total	119		119	

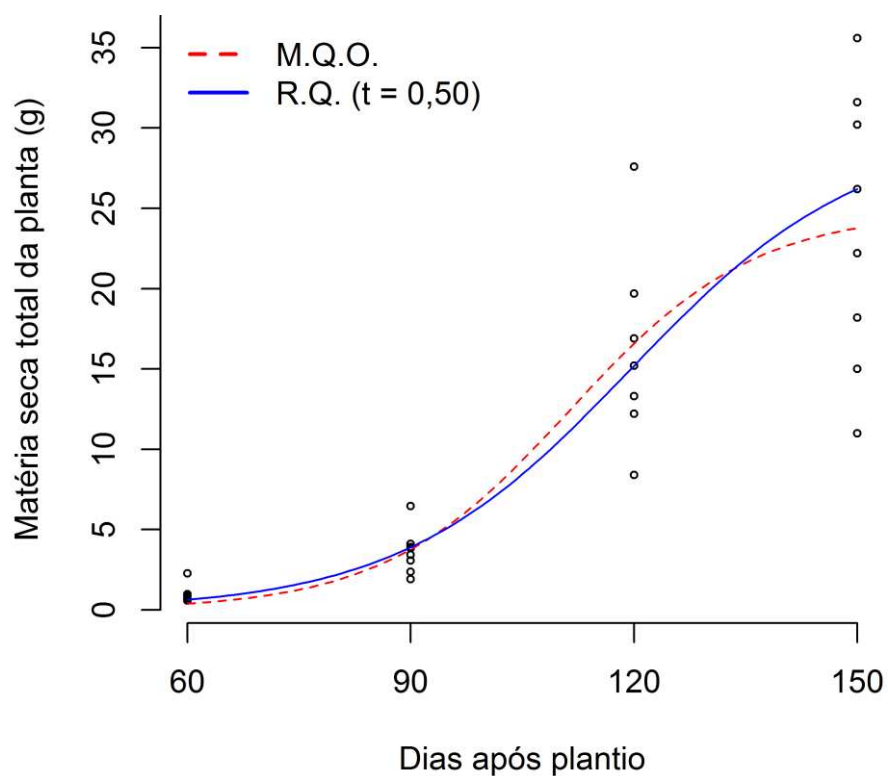


## B – Curvas dos modelos de regressão ajustados

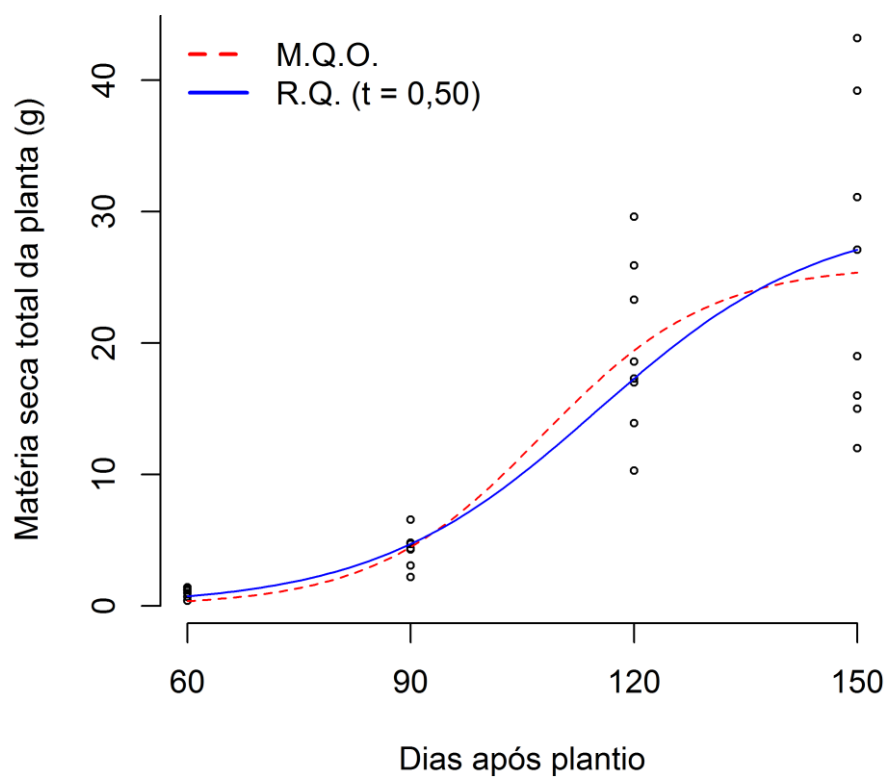
A seguir são apresentadas as curvas dos modelos de regressão não linear ajustadas para cada um dos 30 acessos de alho. Foram ajustados modelos Logísticos de regressão quantílica ( $\tau=0,5$ ) e pelo método dos mínimos quadrados.



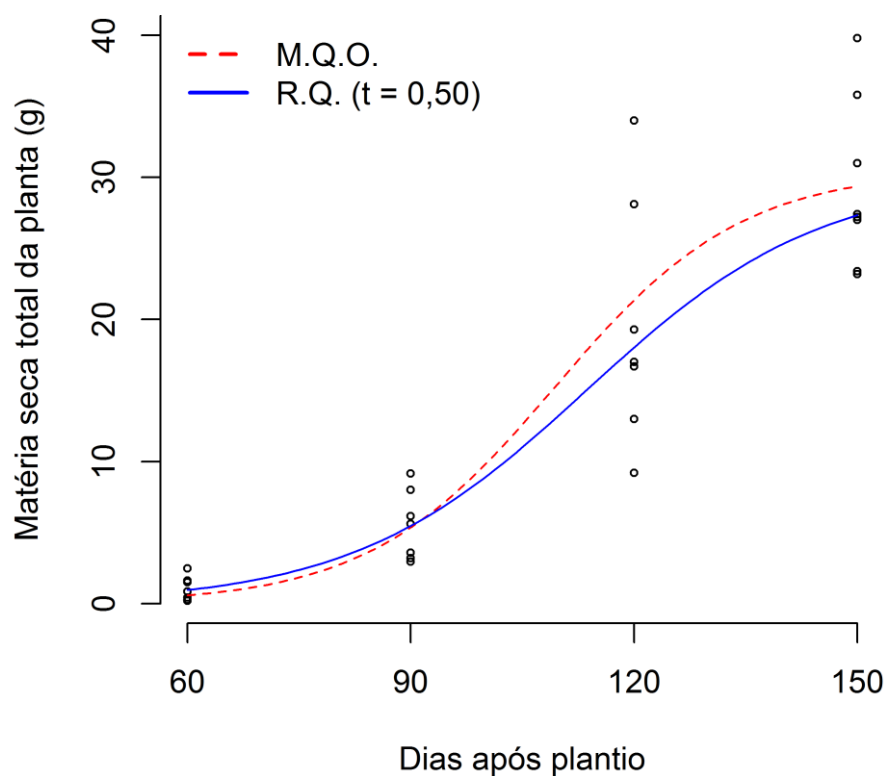
**Figura 1B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 1.



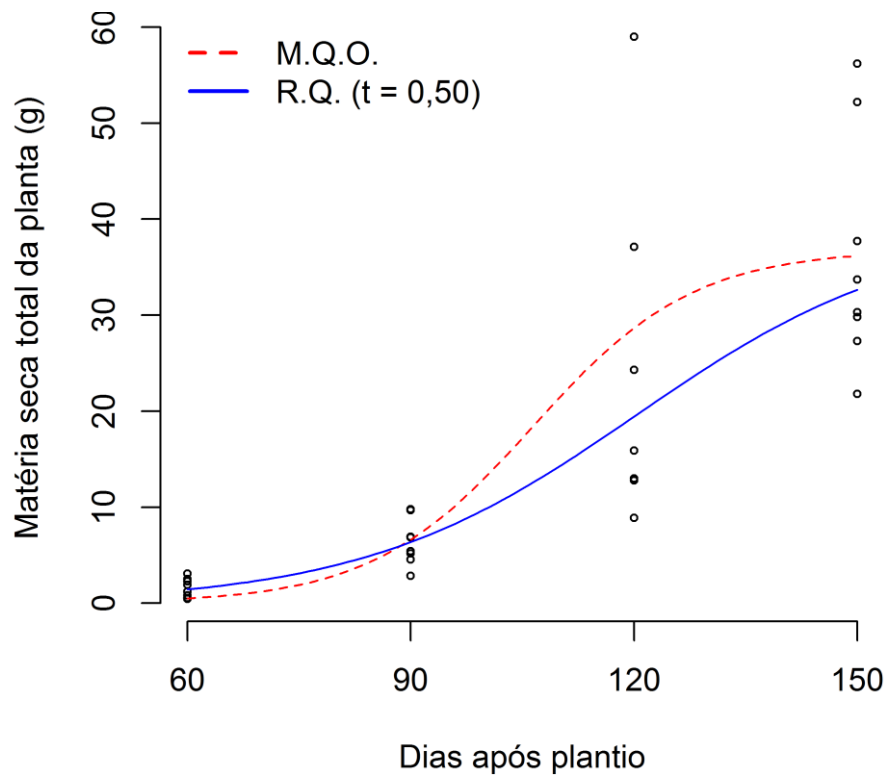
**Figura 2B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 2.



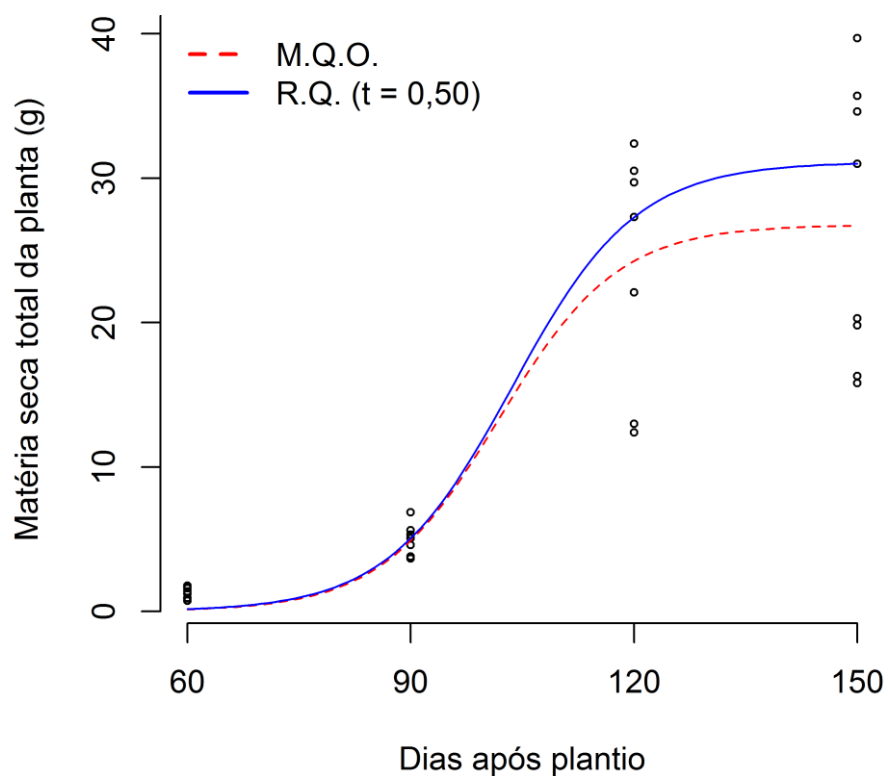
**Figura 3B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 3.



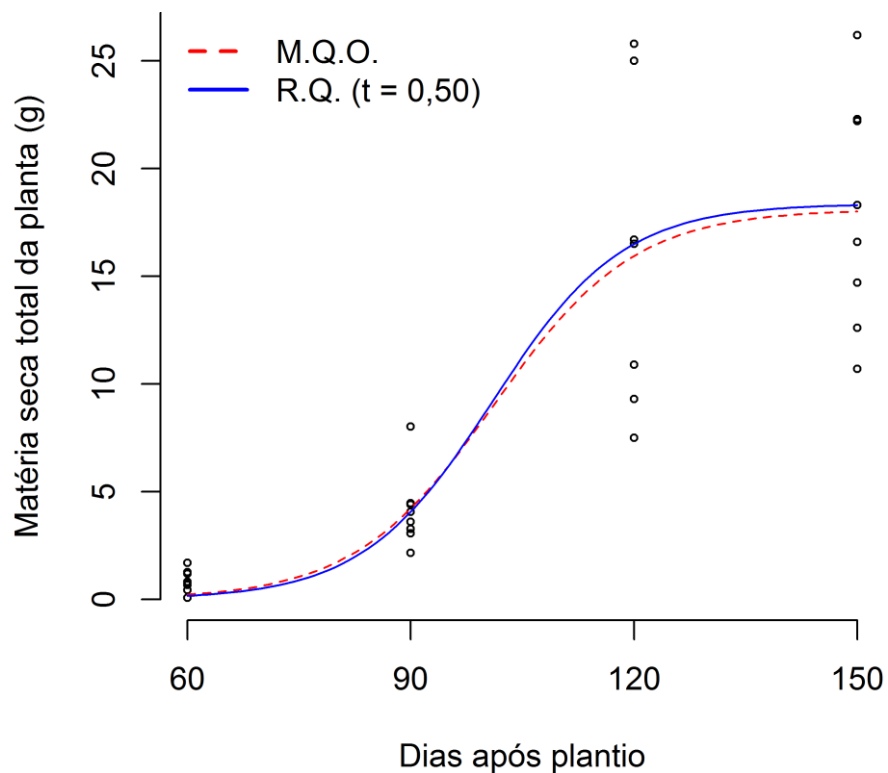
**Figura 4B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 4.



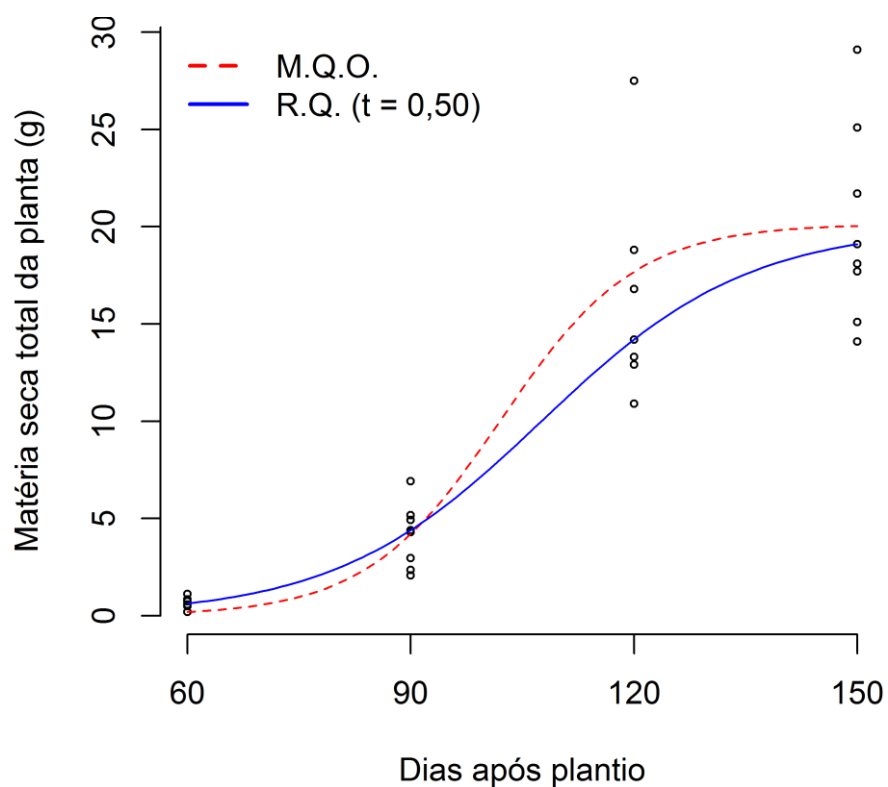
**Figura 5B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 5.



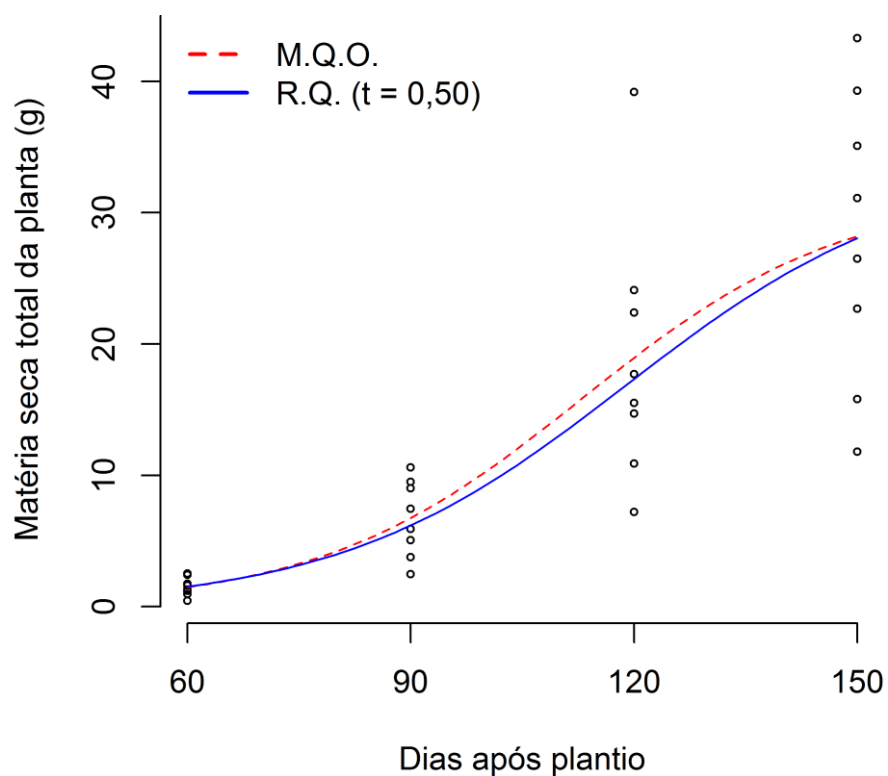
**Figura 6B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 6.



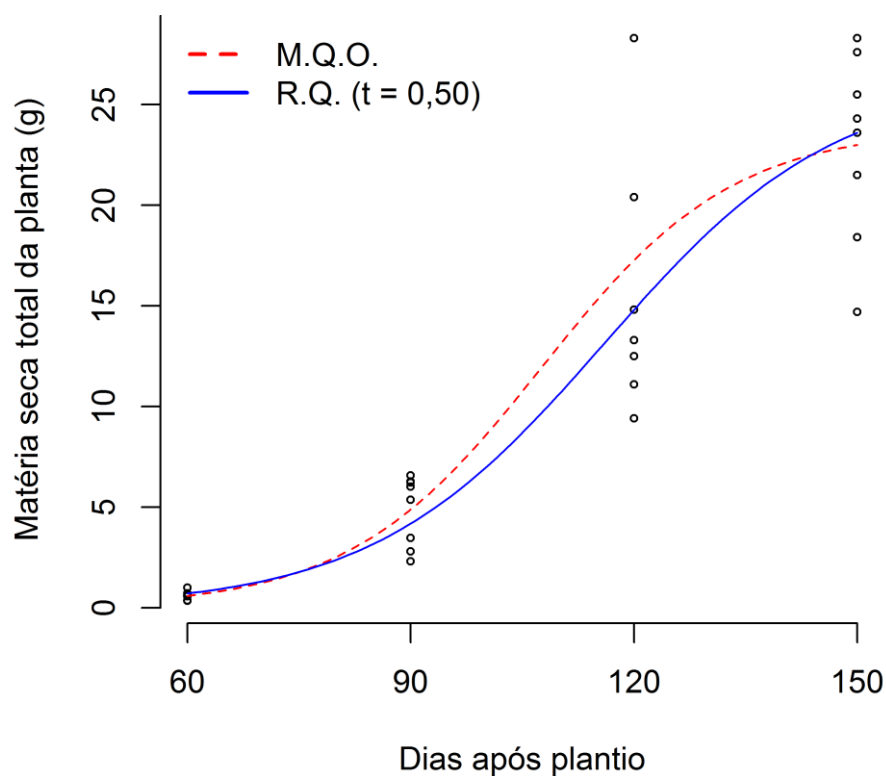
**Figura 7B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 7.



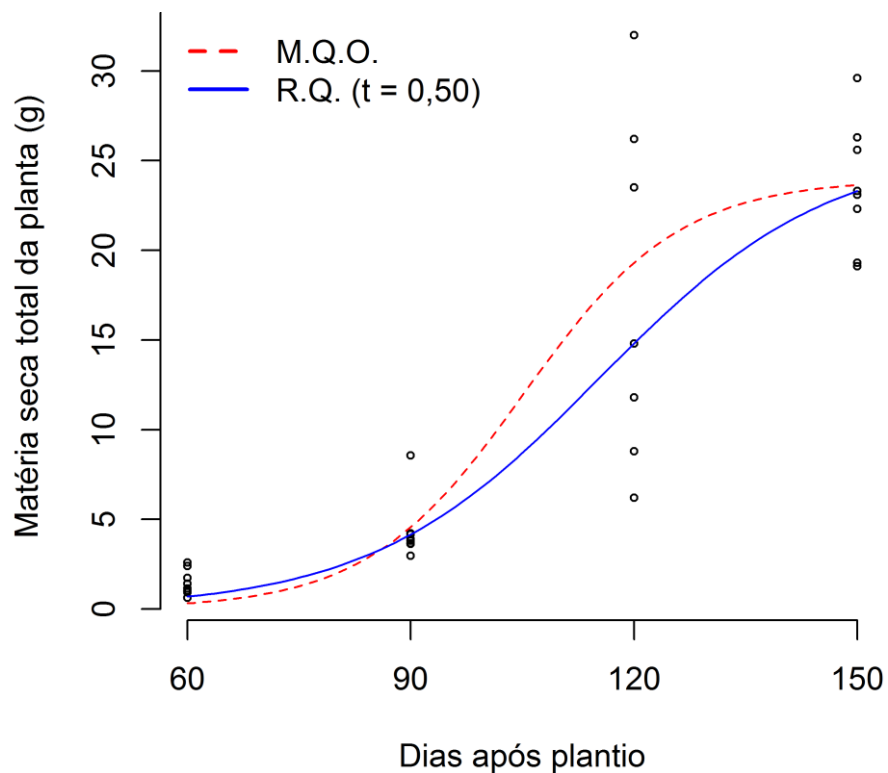
**Figura 8B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 8.



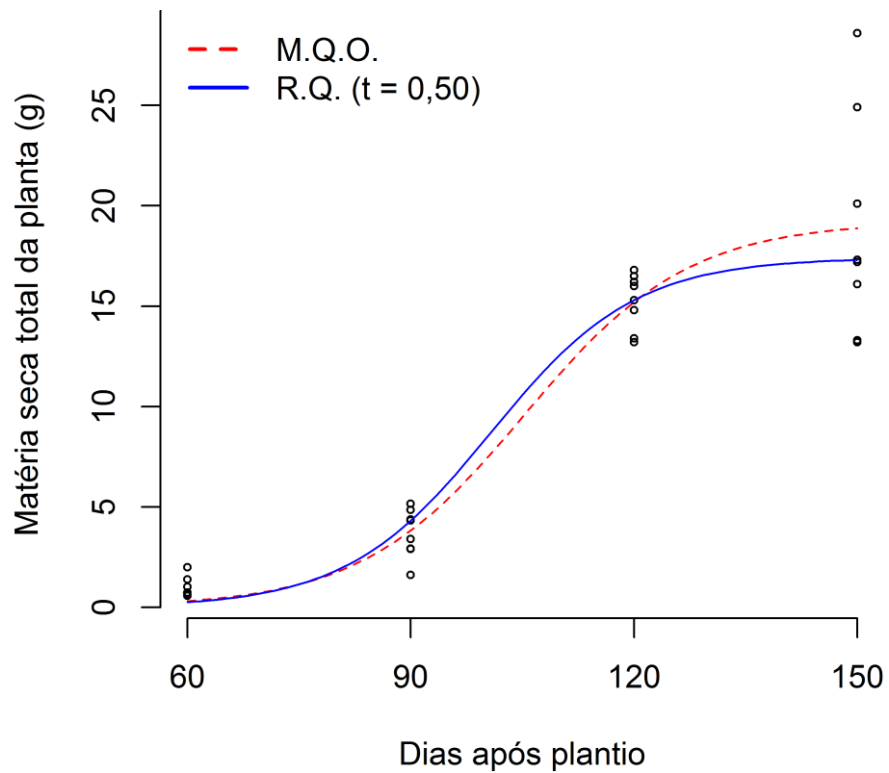
**Figura 9B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 9.



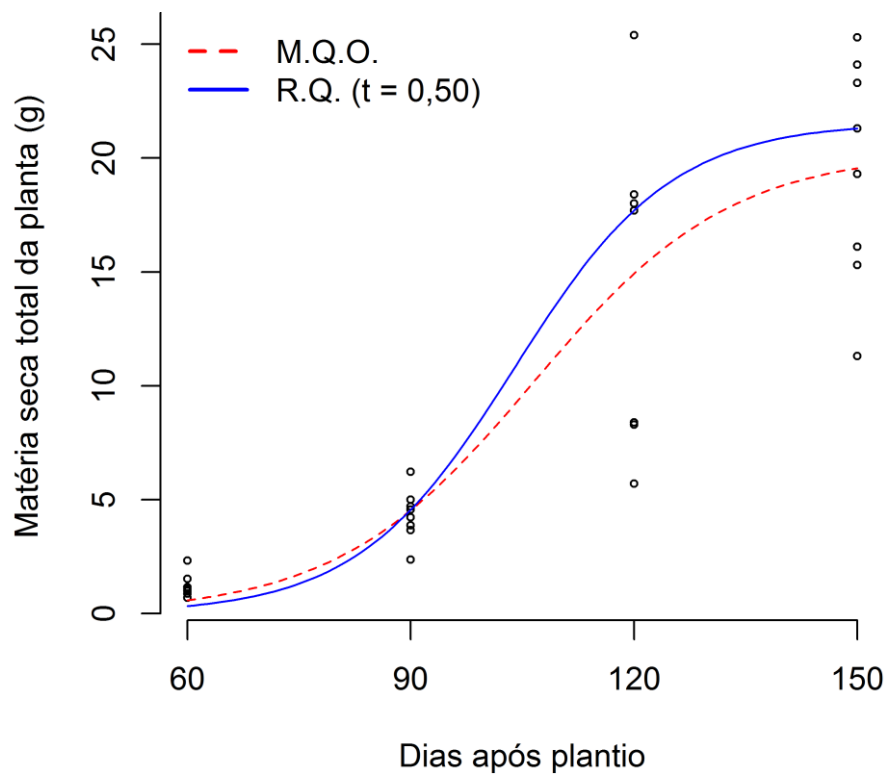
**Figura 10B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 10.



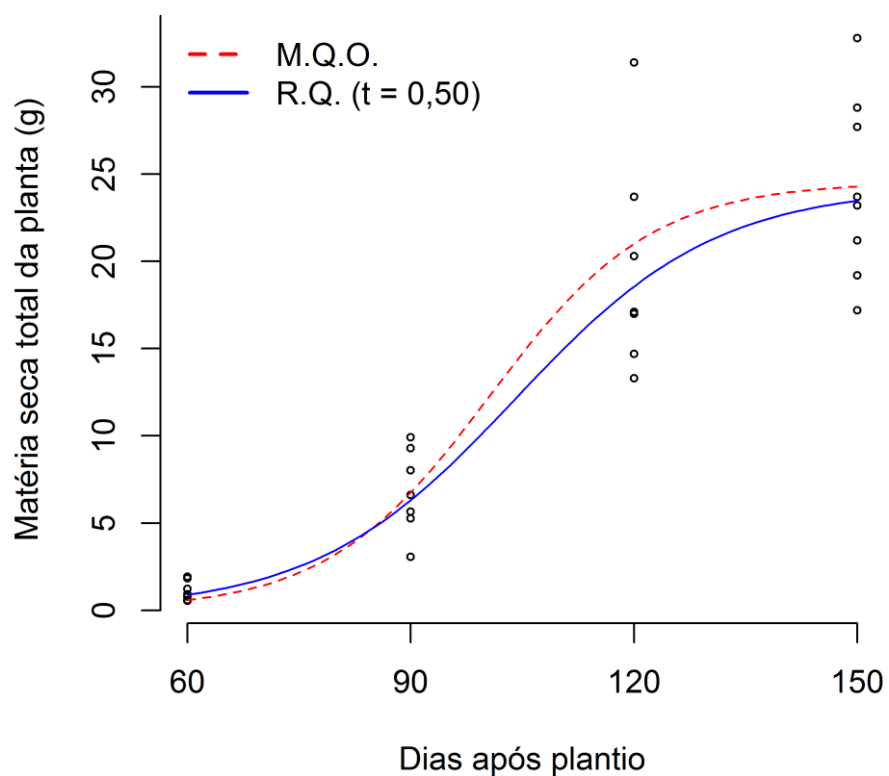
**Figura 11B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 11.



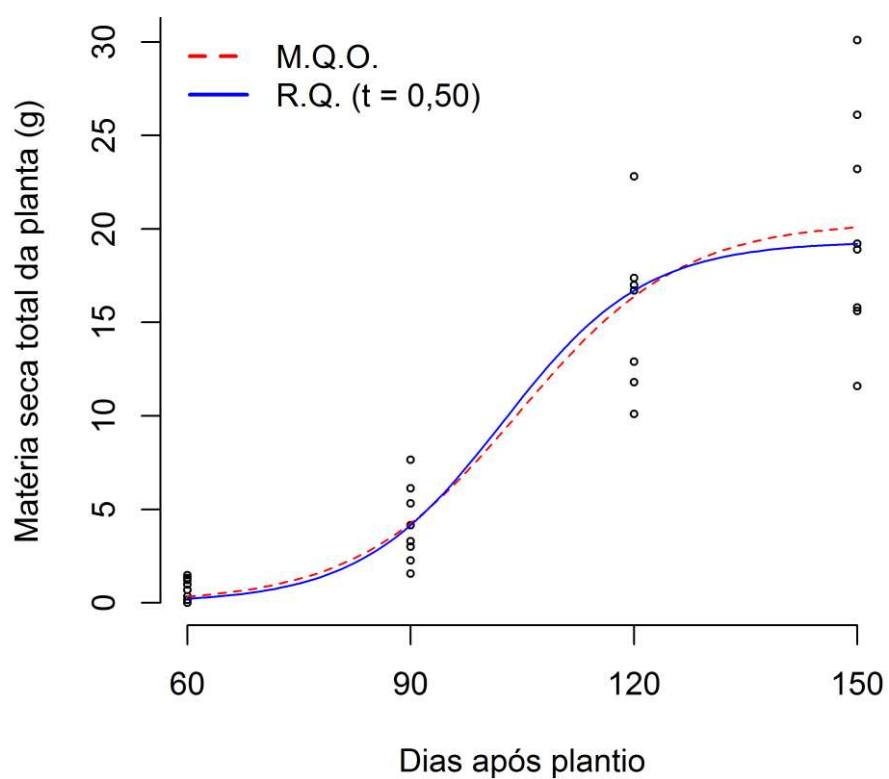
**Figura 12B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 12.



**Figura 13B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 13.

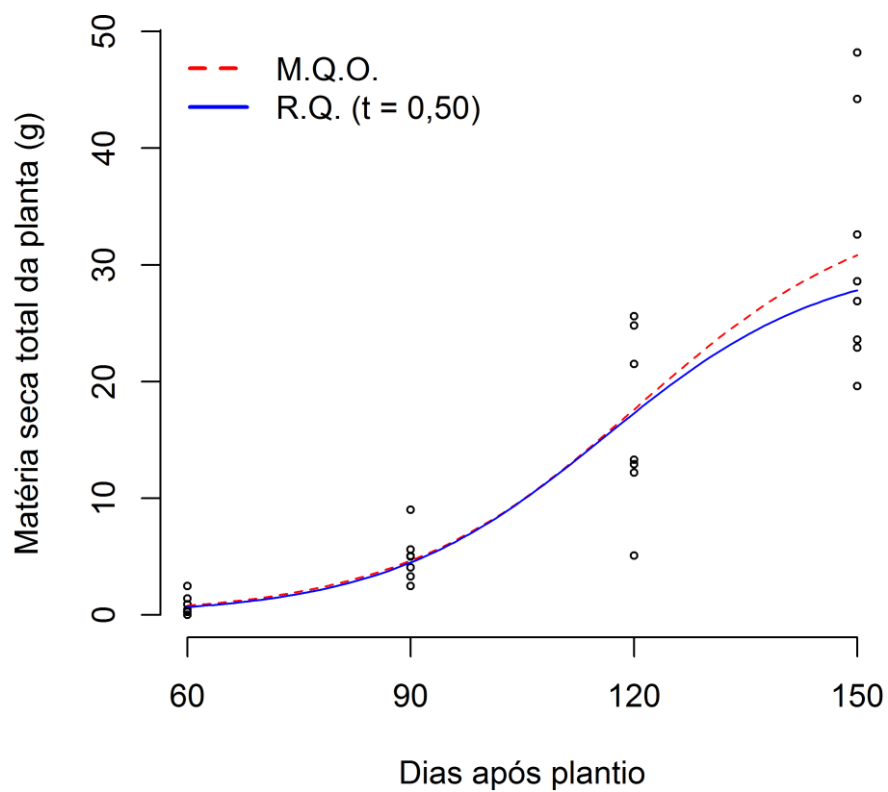


**Figura 14B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 14.

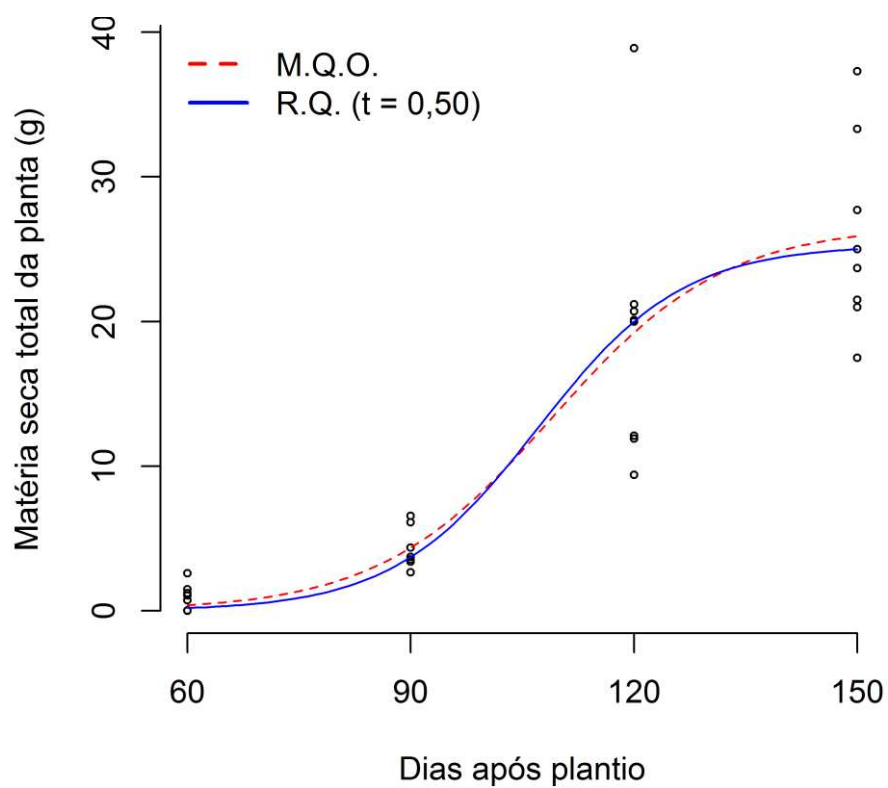


**Figura 15B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 15.

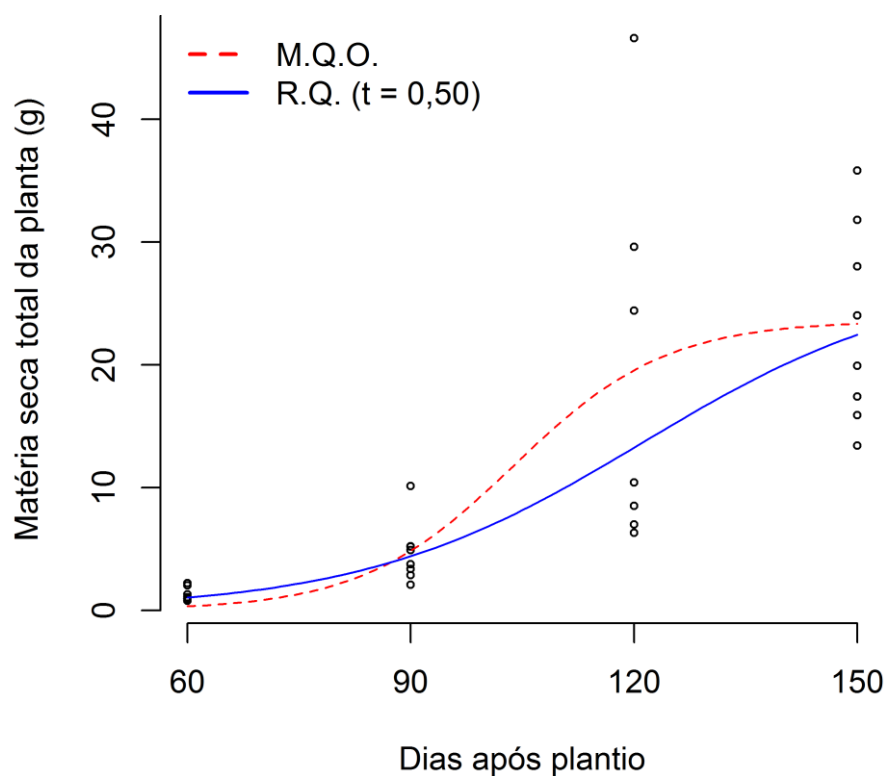




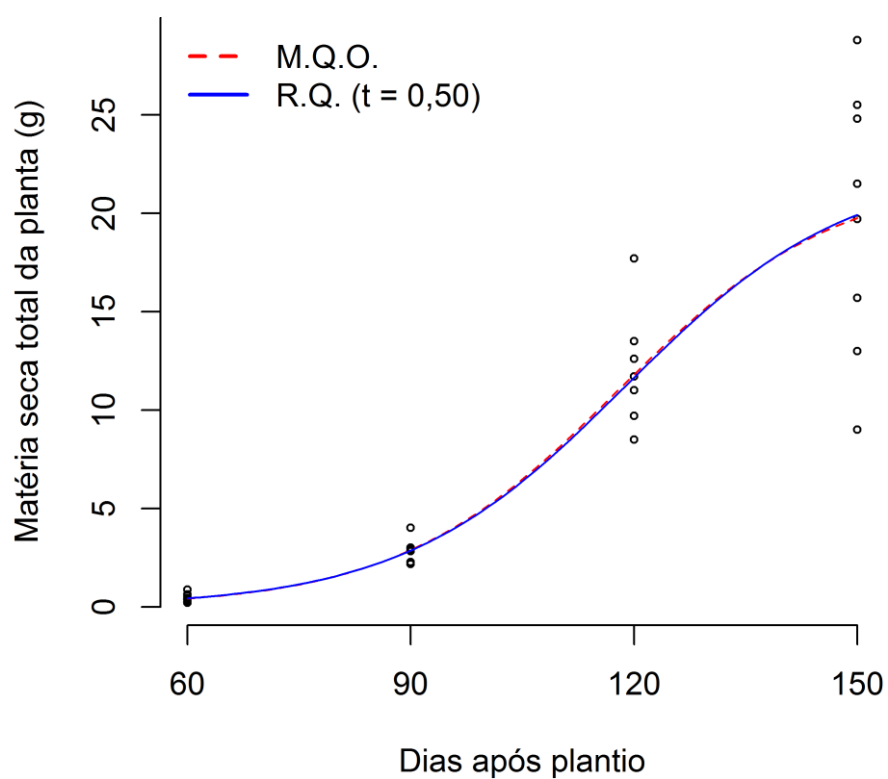
**Figura 16B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 16.



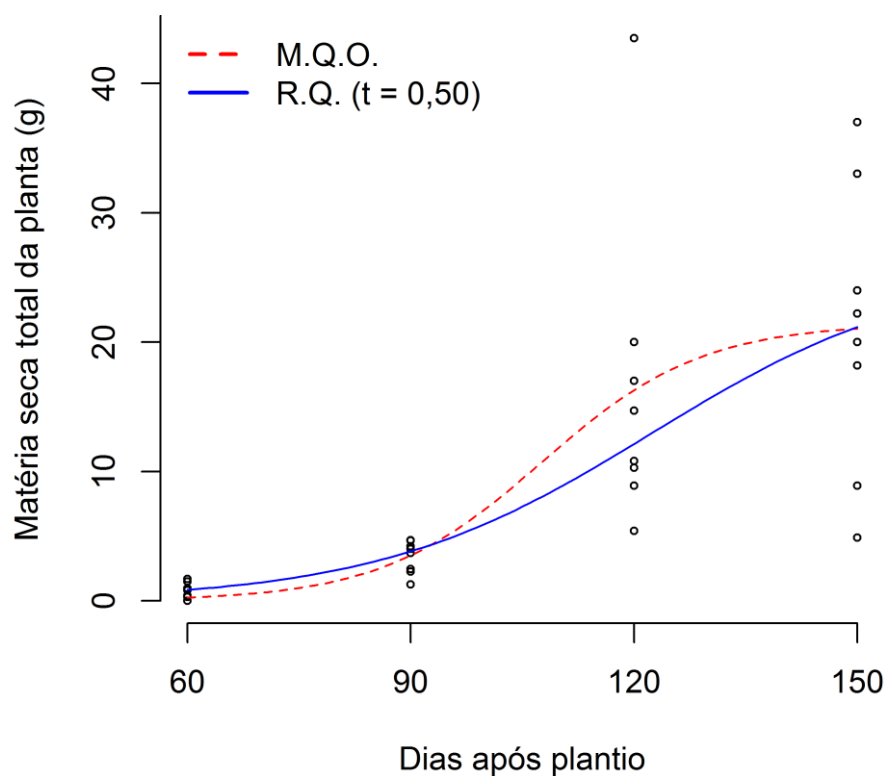
**Figura 17B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 17.



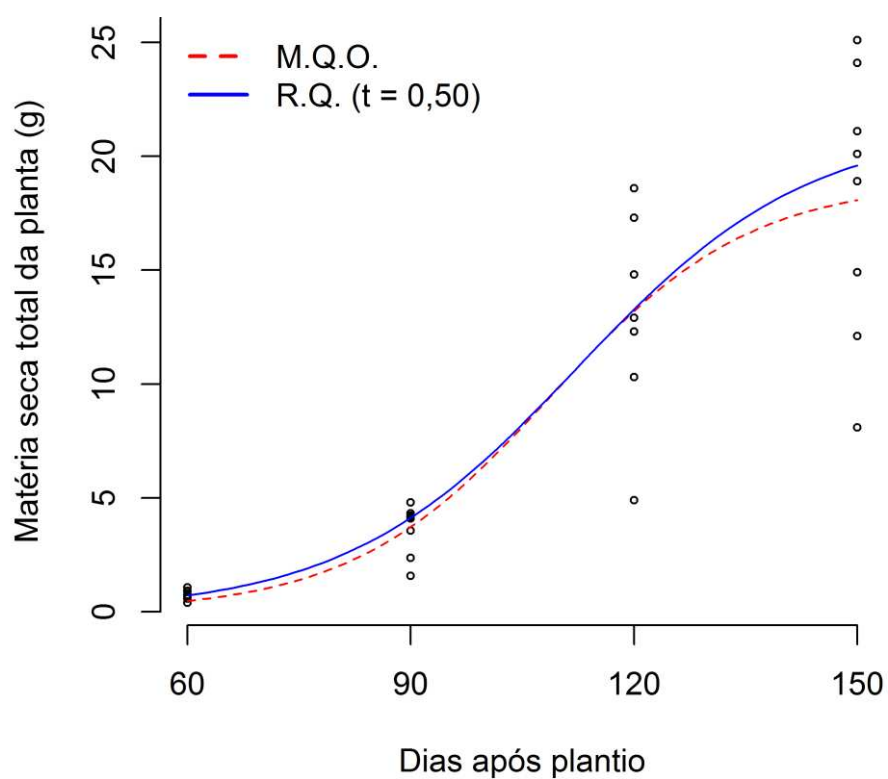
**Figura 18B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 18.



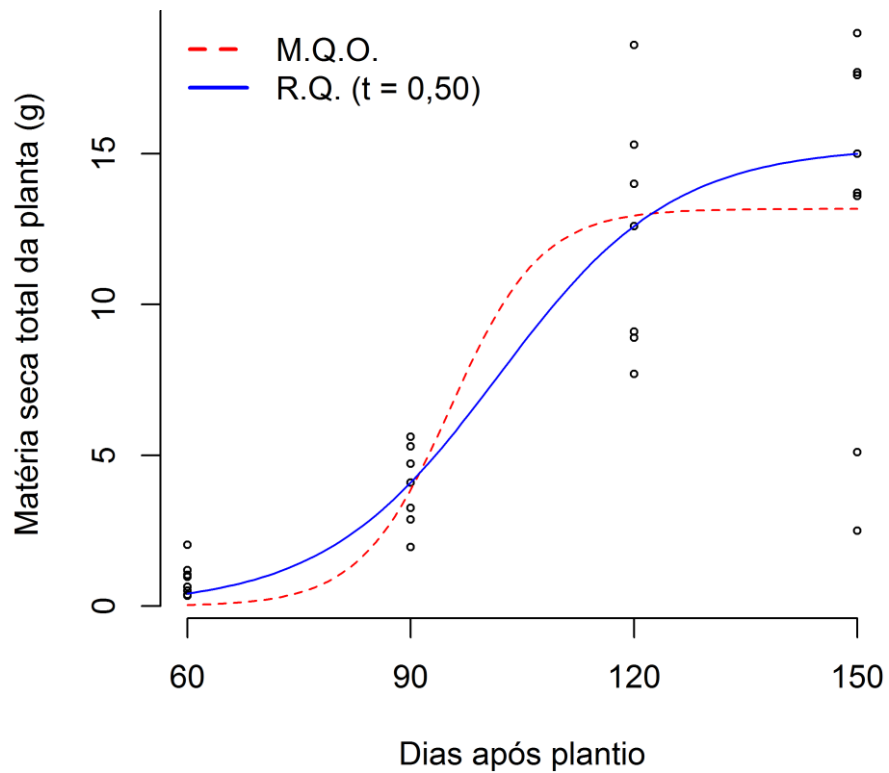
**Figura 19B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 19.



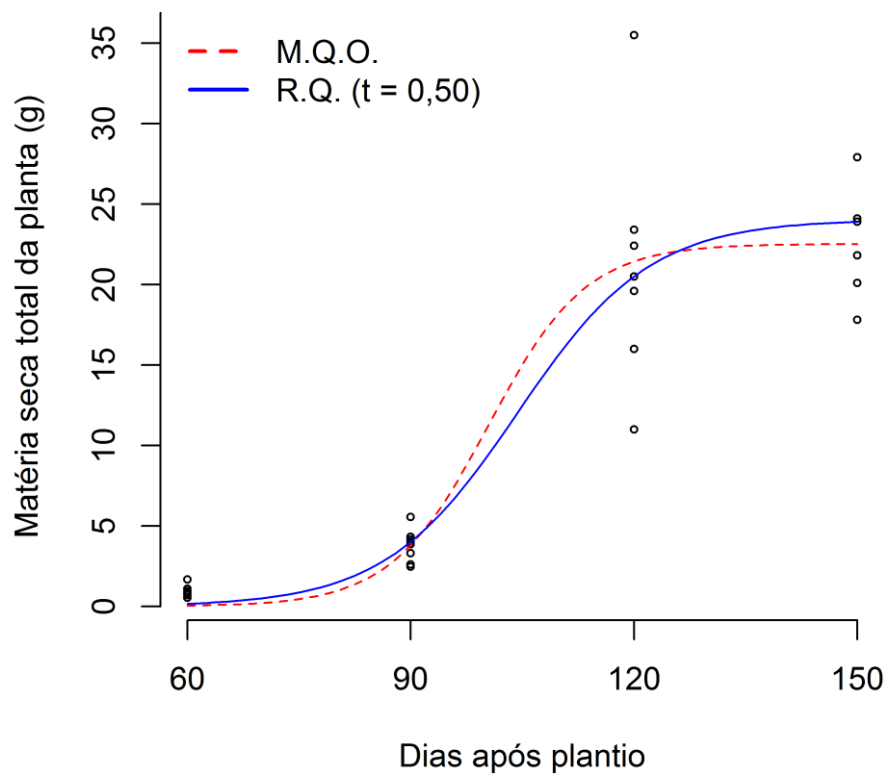
**Figura 20B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 20.



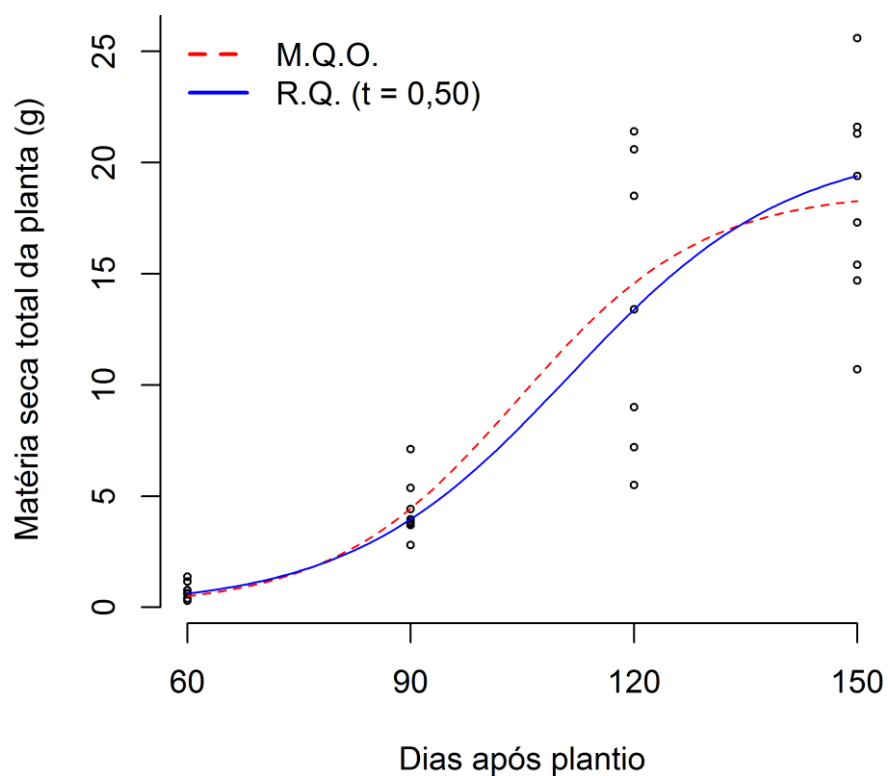
**Figura 21B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 21.



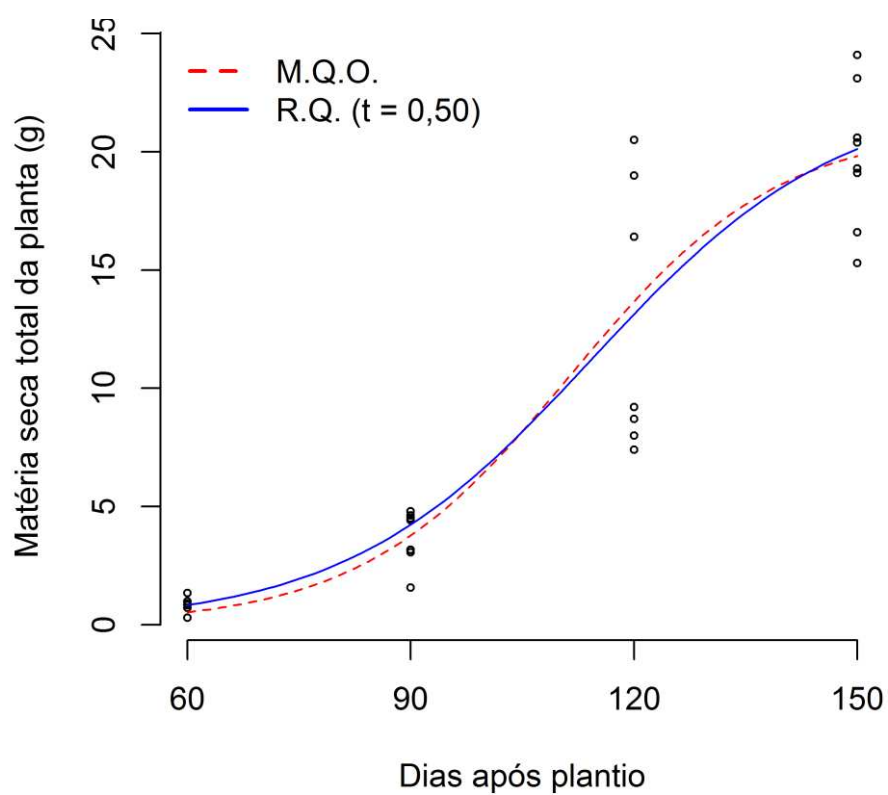
**Figura 22B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 22.



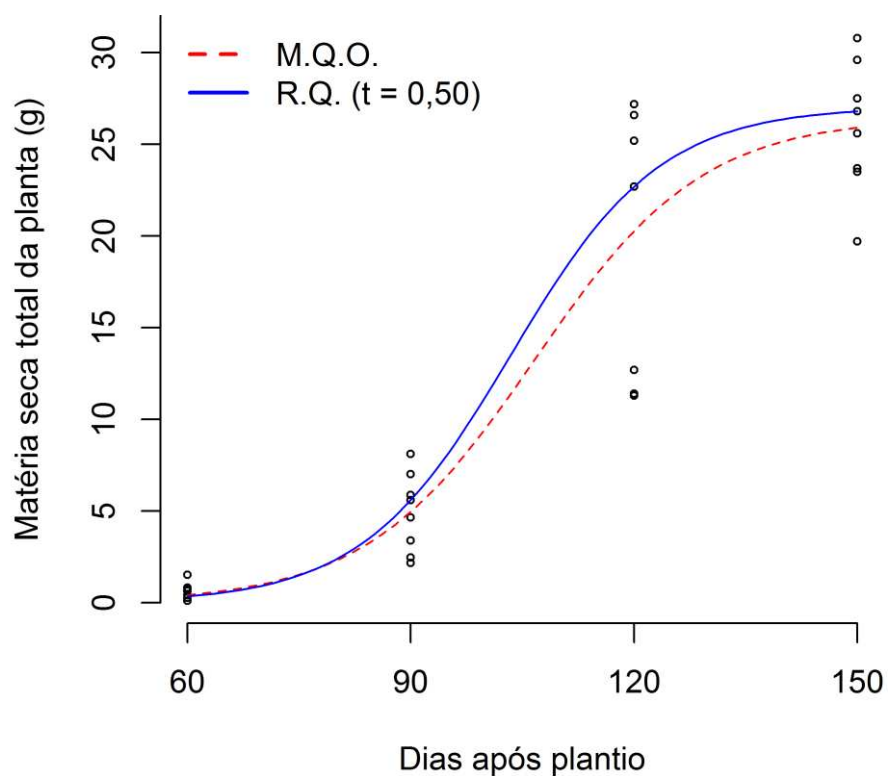
**Figura 23B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 23.



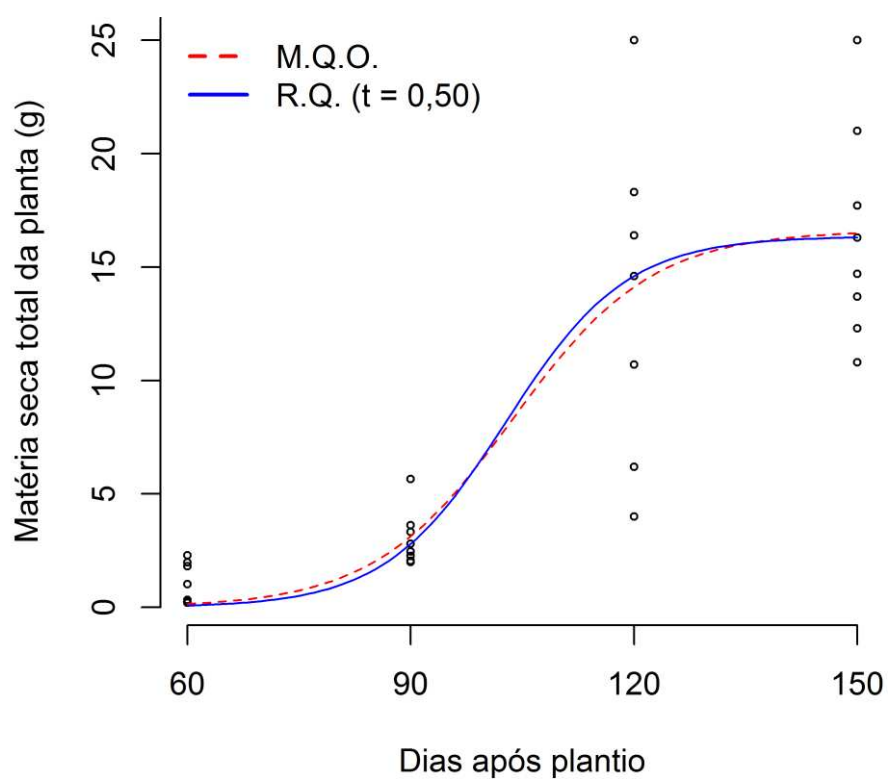
**Figura 24B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 24.



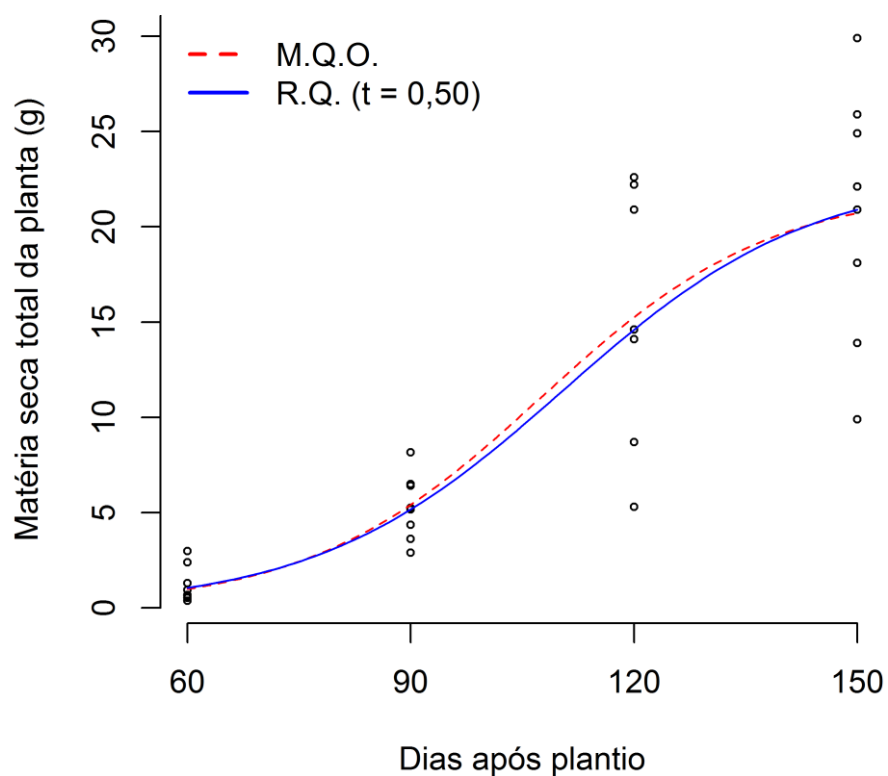
**Figura 25B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 25.



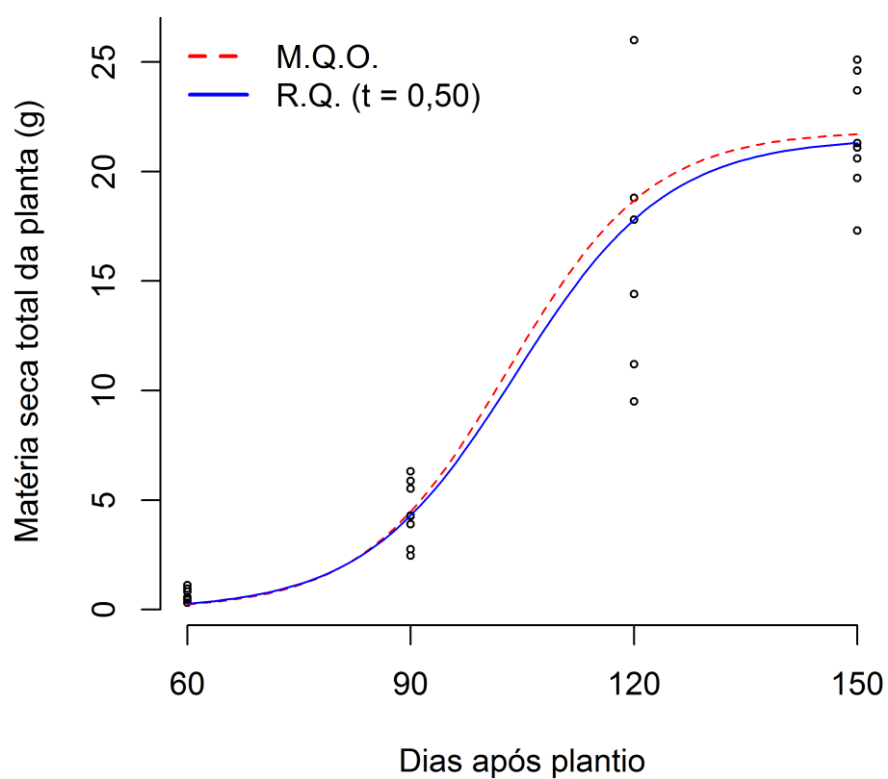
**Figura 26B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 26.



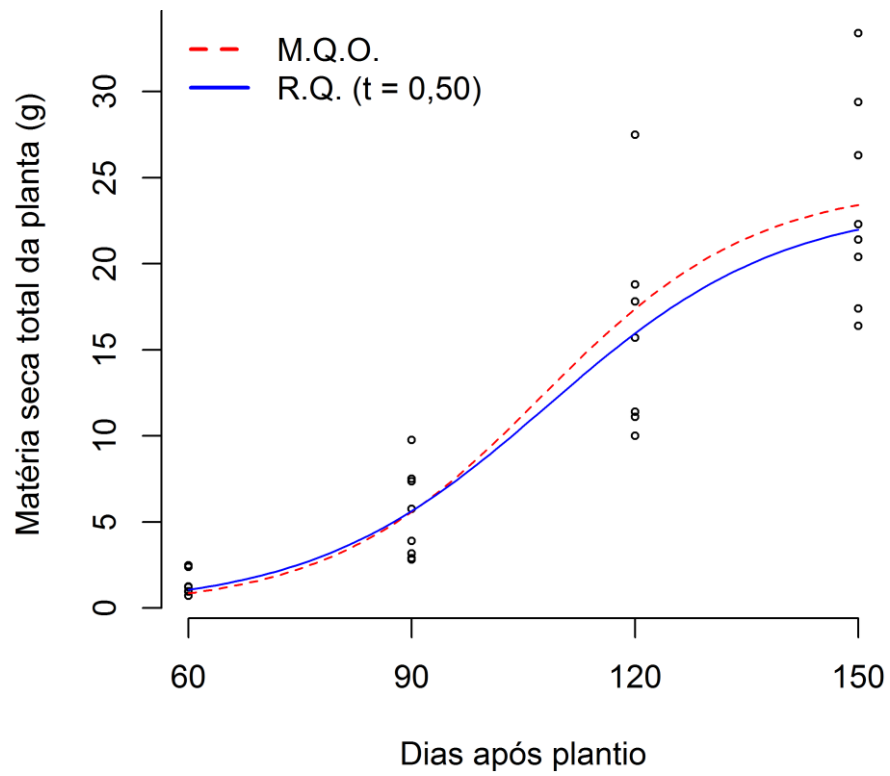
**Figura 27B.**Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 27.



**Figura 28B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 28.



**Figura 29B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 29.

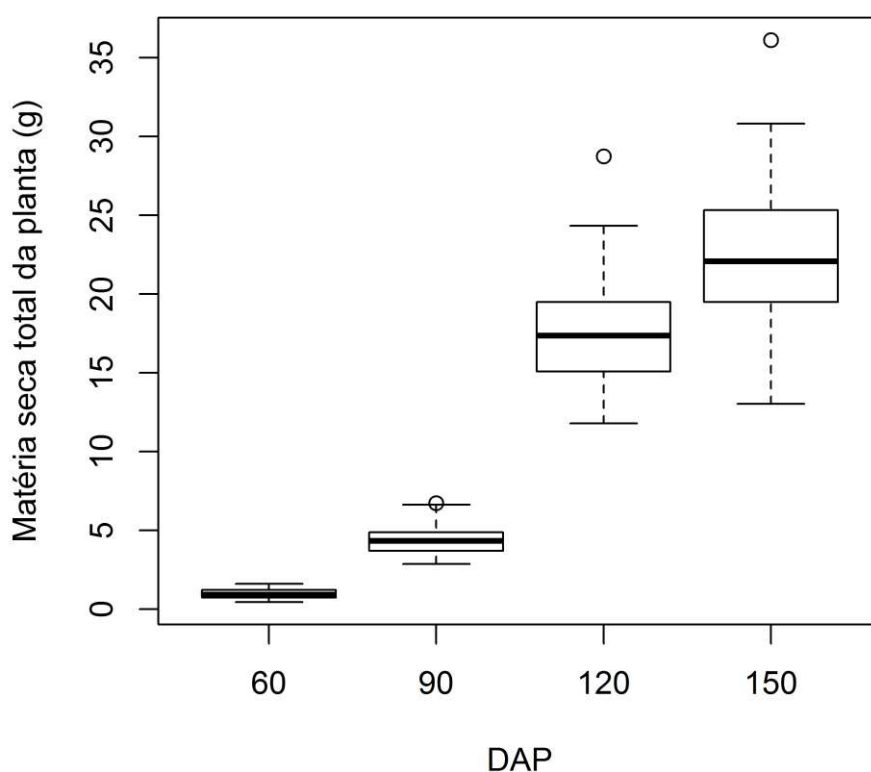


**Figura 30B.** Curvas de acúmulo da matéria seca total da planta (MSTP) ao longo do tempo ajustadas para o acesso de alho 30.

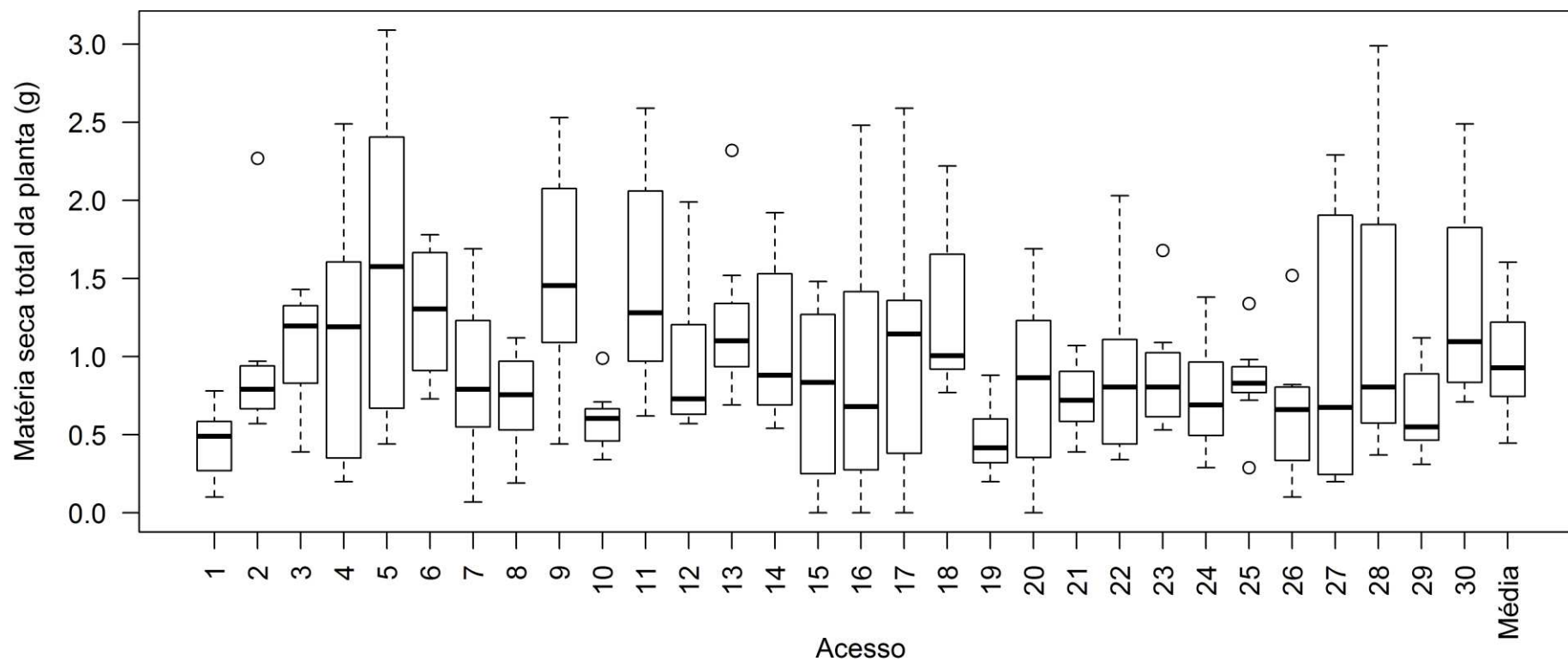


### C – Gráficos de diagrama de caixa (boxplot)

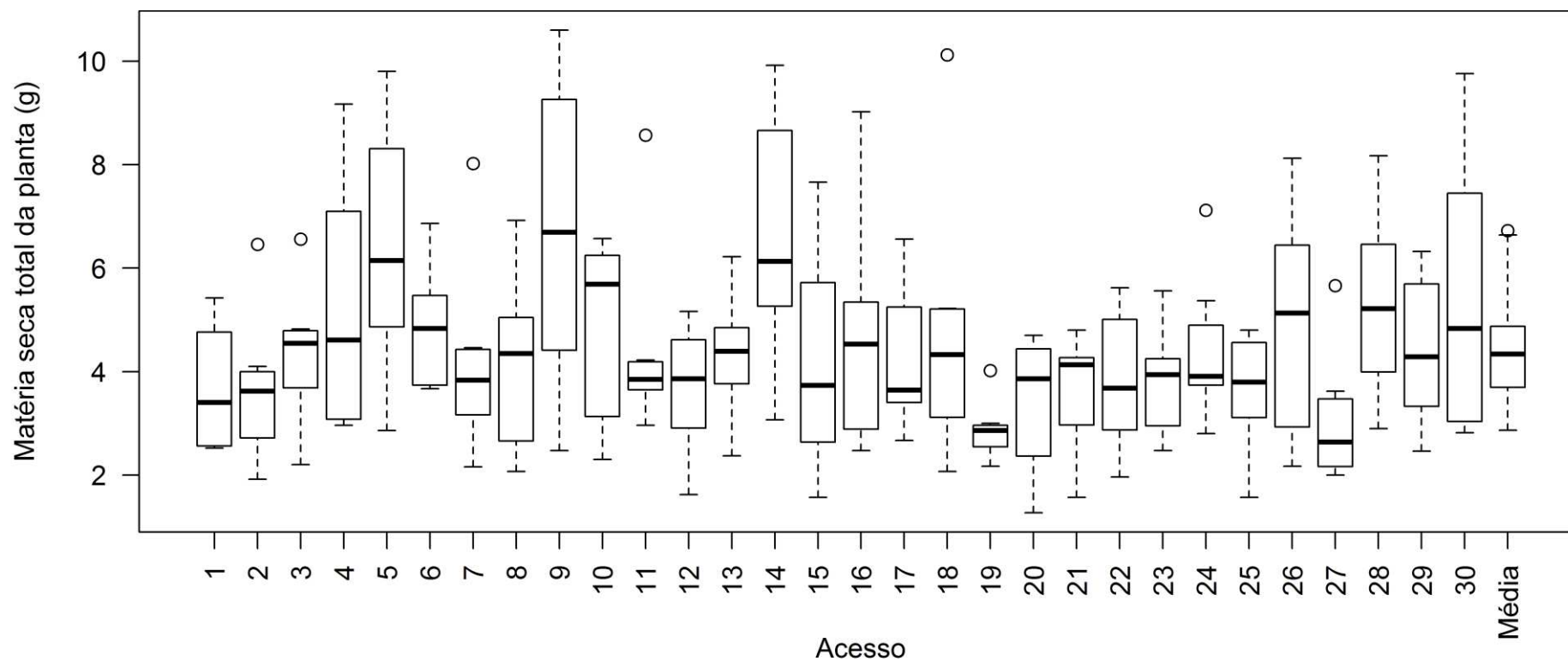
A seguir são apresentados digramas de caixa (*boxplot*) dos dados. Estes gráficos permitem uma análise exploratória dos dados, exibindo informações como assimetria, dispersão, e presença de dados discrepantes. São apresentados diagramas para as médias dos acessos, e para as observações de cada acesso em cada período de avaliação (60, 90, 120 e 150 dias após plantio).



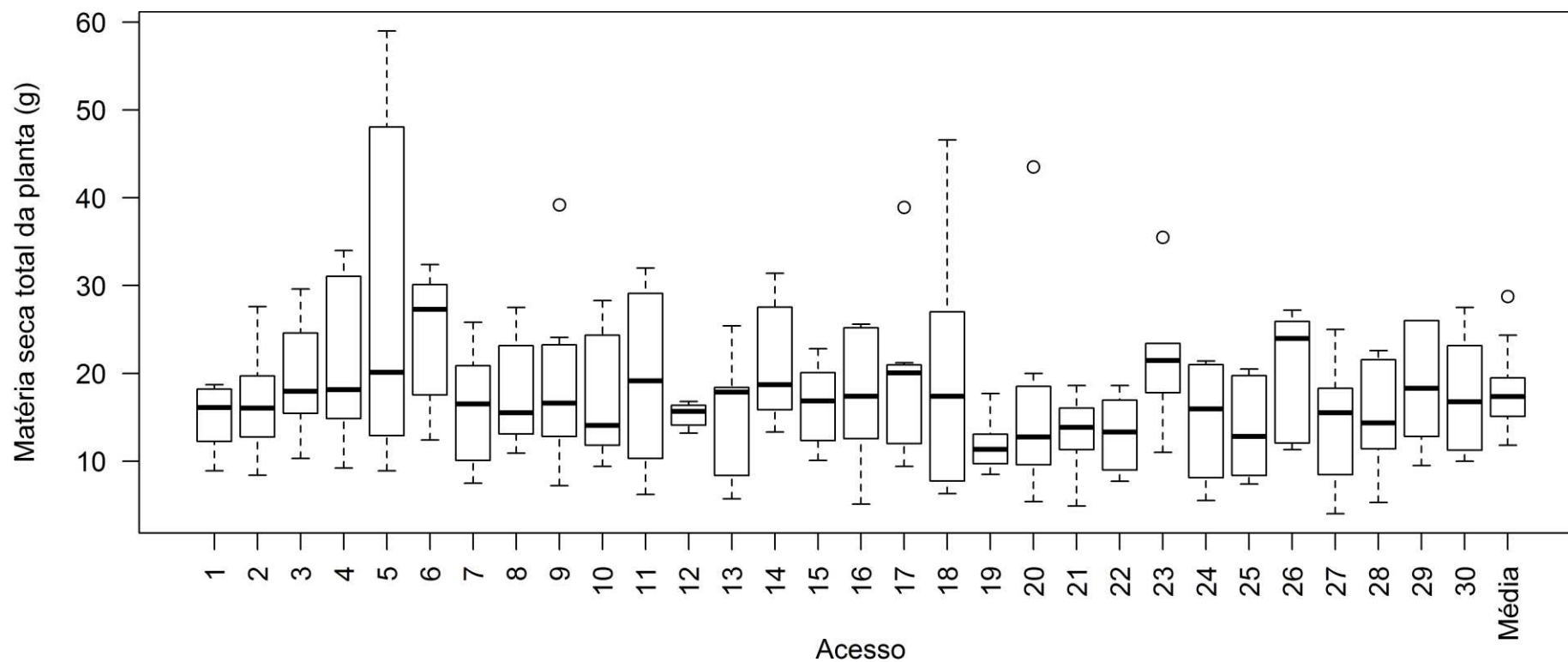
**Figura 1C.** Diagrama de caixas (*boxplot*) do acúmulo de matéria seca total da planta nos quatro períodos para os 30 acessos de alho.



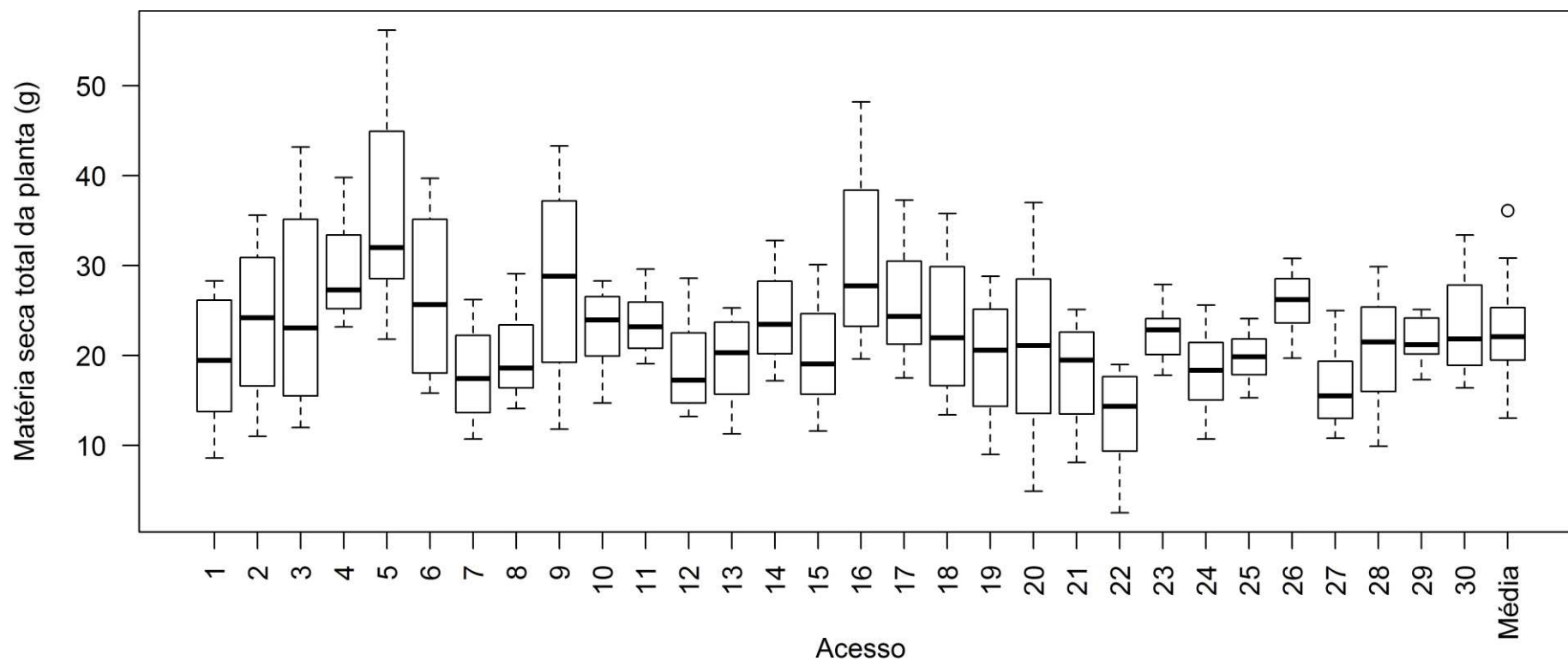
**Figura 2C.** Diagrama de caixas (*boxplot*) do acúmulo de matéria seca total da planta aos 60 dias após plantio para os 30 acessos de alho e para a média destes.



**Figura 3C.** Diagrama de caixas (*boxplot*) do acúmulo de matéria seca total da planta aos 90 dias após plantio para os 30 acessos de alho e para a média destes.



**Figura 4C.** Diagrama de caixas (*boxplot*) do acúmulo de matéria seca total da planta aos 120 dias após plantio para os 30 acessos de alho e para a média destes.



**Figura 5C.** Diagrama de caixas (*boxplot*) do acúmulo de matéria seca total da planta aos 150 dias após plantio para os 30 acessos de alho e para a média destes.

## C – Script das análises no software R

```
### -----  
### Ajustes dos modelos de regressão não linear  
### -----  
  
library(quantreg)  
dados <- read.table("acessos.txt", header = T)  
  
# A respeito dos dados:  
# dados$MSTP = matéria seca total da planta de alho  
# dados$DAP = dias após plantio  
# dados$N = número do acesso  
  
### Ajuste dos modelos de regressão quantílica  
  
# tau = 0,25  
mRQ025 <- nlrq(MSTP~A/(1+B*exp(-C*DAP)), data = dados, tau =  
.25, start = list(A = 22, B = 4000, C = 0.07), trace = TRUE)  
summary(mRQ025)  
  
# tau = 0,50  
mRQ05 <- nlrq(MSTP~A/(1+B*exp(-C*DAP)), data = dados, tau = .5,  
start = list(A = 22, B = 4000, C = 0.07), trace = TRUE)  
summary(mRQ05)  
  
# tau = 0,75  
mRQ075 <- nlrq(MSTP~A/(1+B*exp(-C*DAP)), data = dados, tau =  
.75, start = list(A = 22, B = 4000, C = 0.07), trace = TRUE)  
summary(mRQ075)  
  
### Ajuste do modelo pelo método dos  
### Mínimos Quadrados Ordinários  
mMQO <- nls(MSTP~A/(1+B*exp(-C*DAP)), data = dados, start =  
list(A = 22, B = 4000, C = 0.07))  
summary(mMQO)
```

```

### -----
### Plot das curvas
### -----

plot(dados$DAP, dados$MSTP, cex = .25, type = "n", xlab = "Dias
após plantio", ylab = "Matéria seca total da planta (g)")
points(dados$DAP, dados$MSTP, cex = .5)
curve(coef(mMQO)[1]/(1+coef(mMQO)[2]*exp(-coef(mMQO)[3]*x)), add
= T, lty = 2, col = "red")
curve(coef(mRQ05)[1]/(1+coef(mRQ05)[2]*exp(-coef(mRQ05)[3]*x)),
add = T, lty = 1, col = "blue")
curve(coef(mRQ025)[1]/(1+coef(mRQ025)[2]*exp(-
coef(mRQ025)[3]*x)), add = T, lty = 1, col = "lightslateblue")
curve(coef(mRQ075)[1]/(1+coef(mRQ075)[2]*exp(-
coef(mRQ075)[3]*x)), add = T, lty = 1, col = "midnightblue")
legend("topleft", legend = c("modelo não linear", "tau = 0,75",
"tau = 0,5", "tau = 0,25"), lty = c(2, 1, 1, 1), col = c("red",
"midnightblue", "blue", "lightslateblue"), lwd = 2, bty = "n")

### -----
### Avaliadores de qualidade do ajuste
### -----

# Considerando:
# mMQO = modelo ajustado pelo método dos M. Q. O.
# mRQ = modelo de regressão quantílica (tau = 0,25; 0,5 ou 0,75)
# mod = modelo de regressão quantílica ou de M. Q. O.
# df = graus de liberdade do modelo

### Coeficiente de determinação

# Para modelo ajustado pelo método dos M. Q. O. (R2)
SQR <- summary(mMQO)$sigma^2 * df
SQT <- var(dados$MSTP) * (length(dados$MSTP)-1)
(R2 <- 1-SQR/SQT)

```

```

# Para modelo de regressão quantílica (R1(tau))
SQR <- sum(summary(mRQ)$residuals^2)
SQT <- var(dados$MSTP) * (length(dados$MSTP)-1)
(R1 <- 1-SQR/SQT)

### Quadrado médio do resíduo (QMR)

(QMR <- sum((dados$MSTP - predict(mod, dados$DAP))^2)/(df))

### Desvio médio absoluto do resíduo (DMA)

(DMA <- sum( abs(dados$MSTP - predict(mod, dados$DAP))/
(length(dados$MSTP))))

### Criterio de Akaike (AIC)

AIC(mod)

### -----
### Distâncias entre valores observados dos acessos e preditos
### pelos modelos de R. Q.
### -----

valores <-cbind(dados$N, dados$MSTP, predict(mRQ025,dados$DAP),
predict(mRQ05,dados$DAP), predict(mRQ075,dados$DAP))
colnames(valores) <- c("acesso","obs","rq025","rq05","rq075")
valores # valores observados e preditos

# Distâncias entre valores observados
# e preditos por cada modelo de R. Q.

```



```

dist025 = by(valores[,c(2,3)], valores[,1], function(x)
dist(t(x)))
(dist025<- do.call("rbind", as.list(dist025)))

dist05=by(valores[,c(2,4)], valores[,1], function(x) dist(t(x)))
(dist05<- do.call("rbind", as.list(dist05)))

dist075=by(valores[,c(2,5)], valores[,1], function(x)
dist(t(x)))
(dist075<- do.call("rbind", as.list(dist075)))

(distC <- cbind(unique(valores[,1]),dist025,dist05,dist075))
(dist <- distC[,-c(1)]) # apenas colunas das distâncias

# Menor distância:
minDist <- matrix(NA,nrow(dist),1)
for ( i in 1:nrow(dist)) {
  minDist[i,1] <- which.min(dist[i,])
}

# Classificação dos acessos:
grupos <- cbind(unique(valores[,1]),minDist)
colnames(grupos) <-c("acesso","grupo")
grupos

```