



**ANDREZZA ROSANE SILVA**

**MÉTODOS ESTATÍSTICOS NA ANÁLISE DE DADOS DE PCR  
EM TEMPO REAL**

**LAVRAS – MG**

**2017**

**ANDREZZA ROSANE SILVA**

**MÉTODOS ESTATÍSTICOS NA ANÁLISE DE DADOS DE PCR EM TEMPO REAL**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Dr. Renato Ribeiro de Lima

Orientador

Dr. Antonio Chalfun Junior

Coorientador

**LAVRAS – MG**

**2017**

**Ficha catalográfica elaborada pela Coordenadoria de Processos Técnicos  
da Biblioteca Universitária da UFLA**

Silva, Andrezza Rosane

Métodos estatísticos na análise de dados de PCR em tempo real / Andrezza Rosane Silva. – Lavras : UFLA, 2017.

78 p. : il.

Dissertação(mestrado acadêmico)–Universidade Federal de Lavras, 2017.

Orientador: Dr. Renato Ribeiro de Lima.

Bibliografia.

1. qPCR. 2. Método C<sub>q</sub> comparativo. 3. Modelo de eficiência calibrada. 4. ANOVA. 5. Modelos mistos. I. Universidade Federal de Lavras. II. Título.

**ANDREZZA ROSANE SILVA**

**MÉTODOS ESTATÍSTICOS NA ANÁLISE DE DADOS DE PCR EM TEMPO REAL**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 25 de Maio de 2017.

Profa. Dra. Izabela Regina Cardoso Oliveira UFLA  
Prof. Dr. Horllys Gomes Barreto UFT

Dr. Renato Ribeiro de Lima  
Orientador

Dr. Antonio Chalfun Junior  
Coorientador

**LAVRAS – MG**  
**2017**

*À minha mãe Milva, por ser meu maior exemplo.  
Dedico*

## AGRADECIMENTOS

Agradeço a Deus e a Nossa Senhora Aparecida, por guiar meus passos e não me deixar desistir.

À minha mãe Milva, pelo amor, amizade, carinho, dedicação e incentivo que me proporcionou.

À minha irmã Ana Luíza, pelo companheirismo, amizade e ajuda de sempre.

À minha família, pelo apoio, em especial ao meu avô Jadyr, que sempre se mostrou um exemplo a ser seguido.

Ao Luís, pela ajuda, paciência, amor e incentivo.

Às amigas Érica e Bruna, pela convivência, amizade, apoio e confiança.

Às amigas de Itapecerica pela confiança e amizade.

Aos amigos e colegas da Pós-graduação em Estatística e Experimentação Agropecuária, em especial, aos amigos do mestrado 2015/1, pela convivência e aprendizados obtidos.

Ao professor e orientador Renato Ribeiro de Lima, pelos incentivos, correções, paciência e ensinamentos que tornaram possíveis a realização deste trabalho.

Ao professor Antonio Chalfun Junior e ao André Almeida Lima pelos ensinamentos e disponibilização dos dados utilizados neste trabalho.

Aos professores Izabela Regina Cardoso Oliveira e Horllys Gomes Barreto pela participação da banca de defesa.

À Universidade Federal de Lavras, pela oportunidade e estrutura oferecida, especialmente aos professores e funcionários do Departamento de Estatística.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela concessão da bolsa.

## RESUMO

A qPCR é uma técnica amplamente utilizada em experimentos de quantificação da expressão gênica. No entanto, considerações estatísticas como testes de significância, valores-p e intervalos de confiança não são utilizados em muitos trabalhos que envolvem a técnica de qPCR. Assim, pesquisadores que não realizam a análise estatística de forma adequada podem obter inferências equivocadas. O objetivo deste trabalho é apresentar e comparar diferentes métodos de análise de dados de expressão gênica, obtidos utilizando-se a técnica de qPCR. Nesta comparação foram utilizados dados de um experimento para avaliar a expressão gênica em mudas de café (*Coffea arabica*) obtidas em diferentes condições de estresse hídrico. Os métodos aplicados e comparados foram: modelo de eficiência calibrada, método  $C_q$  comparativo, análise de variância considerando modelos fixos e mistos. Para fins de comparação, para cada método, foram obtidos os valores referentes às razões da expressão relativa e seus respectivos intervalos de confiança a 95%. As análises foram realizadas utilizando os softwares R e SAS. Os resultados indicaram que os valores obtidos para a razão utilizando os métodos  $C_q$  comparativo, modelo de eficiência calibrada, modelos fixo e misto, foram em geral similares. Os resultados obtidos considerando o modelo misto foram melhores por apresentar, na maioria dos casos, intervalos de confiança com menores amplitudes e, conseqüentemente, maior precisão nas inferências realizadas.

**Palavras-chave:** qPCR. Método  $C_q$  comparativo. Modelo de eficiência calibrada. ANOVA. Modelos mistos.

## ABSTRACT

The qPCR is a widely technique used in the experiments related to gene expression quantification. However, in several cases, statistical formalism like significance tests, p-value and confidence intervals are not used in the data analysis from experiments which involve the qPCR technique. Thus, lack of statistical formalism leads to invalid inferences. The purpose of this work is to present and to compare different methods to analyse gene expression data obtained by using the qPCR. Data from an experiment with coffee plants under water-deficit conditions were used to evaluate the methods. The methods which were evaluate in this study were: efficiency calibrated model, comparative  $C_q$  method, analysis of variance considering fixed and mixed models. The ratios, which represent the relative expressions, with the respective confidence intervals were obtained and used to compare different methods. The analyses had been done using the softwares R and SAS. The results indicated that the ratio values obtained by the comparative  $C_q$  method, efficiency calibrated model, fixed and mixed models are close. However, the best results were obtained using the mixed models, because the width of de confidence intervals were the smallest. Thus, the precision of the estimated effects was better than the estimates obtained by the other methods.

**Keywords:** qPCR. Comparative  $C_q$  method. Efficiency calibrated model. ANOVA. Mixed models.



## LISTA DE FIGURAS

Figura 2.1 – Curvas de qPCR . . . . .	14
Figura 2.2 – Quantificação por qPCR . . . . .	17
Figura 4.1 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas pelo modelo de eficiência calibrada para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões . . . . .	39
Figura 4.2 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas pelo método $C_q$ comparativo para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões . . . . .	43
Figura 4.3 – Estimação do parâmetro $\lambda$ pelo método de Box-Cox para o modelo fixo de ANOVA utilizando o $C_q$ médio . . . . .	44
Figura 4.4 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas utilizando os valores das estimativas de $\Delta\Delta C_q$ obtidas pelo modelo fixo com estrutura de matriz de variâncias e covariâncias ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões . . . . .	48
Figura 4.5 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas utilizando os valores das estimativas de $\Delta\Delta C_q$ obtidas pelo modelo misto com estrutura de matriz de variâncias e covariâncias ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões . . . . .	51

Figura 4.6 – Comparação das razões da expressão relativa do gene CaACS1-like obtidas utilizando os seguintes métodos: modelo de eficiência calibrada; método  $C_q$  comparativo; modelo fixo sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1) e modelo misto sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1). Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões . . . . . 55

## LISTA DE TABELAS

Tabela 4.1 – Estimativas das razões da expressão relativa do gene CaACS1-like obtidas pelo modelo de eficiência calibrada para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada razão são apresentados o erro padrão, o valor-p do teste de Shapiro-Wilk e o intervalo de confiança a 95% . . . . .	38
Tabela 4.2 – Estimativas de $\Delta\Delta C_q$ obtidas pelo método $C_q$ comparativo para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa de $\Delta\Delta C_q$ são apresentados erro padrão, o valor-p do teste de Shapiro-Wilk, o valor-p do teste t de Student e o intervalo de confiança a 95% . . . . .	41
Tabela 4.3 – Estimativas das razões da expressão relativa do gene CaACS1-like obtidas pelo método $C_q$ comparativo para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa da razão são apresentados o erro padrão, o valor-p do teste de Shapiro-Wilk e o intervalo de confiança a 95% . . . . .	41
Tabela 4.4 – Resumo da análise de variância para o modelo fixo utilizando como valor de $C_q$ a média das réplicas após transformação Box-Cox . . . . .	45
Tabela 4.5 – Resumo da análise de variância para o modelo fixo com a estrutura de matriz de variâncias e covariâncias residuais do tipo ARH(1) para o conjunto de dados utilizando como valor de $C_q$ todas as réplicas . . . . .	46
Tabela 4.6 – Estimativas de $\Delta\Delta C_q$ obtidas pelo modelo fixo com estrutura de matriz de variâncias e covariâncias residuais ARH(1) ajustado utilizando o $C_q$ referente a todas as réplicas, para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa de $\Delta\Delta C_q$ são apresentados o erro padrão, o valor-p do teste t de Student e o intervalo de confiança a 95% . . . . .	46

Tabela 4.7 – Estimativas das razões da expressão relativa do gene CaACS1-like utilizando os valores de $\Delta\Delta C_q$ estimados pelo modelo fixo com estrutura de matriz de variâncias e covariâncias residuais ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa da razão são apresentados o intervalo de confiança a 95% . . . . .	47
Tabela 4.8 – Resumo da análise de variância para o modelo misto com a estrutura de matriz de variâncias e covariâncias residuais do tipo ARH(1) . . . . .	49
Tabela 4.9 – Estimativas de $\Delta\Delta C_q$ obtidas pelo modelo misto com estrutura de matriz de variâncias e covariâncias residuais ARH(1), para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa de $\Delta\Delta C_q$ são apresentados o erro padrão, o valor-p do teste t de Student e o intervalo de confiança a 95% . . . . .	49
Tabela 4.10 – Estimativas das razões da expressão relativa do gene CaACS1-like utilizando os valores de $\Delta\Delta C_q$ estimados pelo modelo misto com estrutura de matriz de variâncias e covariâncias residuais ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa da razão são apresentados o intervalo de confiança a 95% . . . . .	50
Tabela 4.11 – Comparação dos resultados obtidos para as razões e respectivos intervalos de confiança a 95% e amplitudes utilizando o modelo de eficiência calibrada, o método $C_q$ comparativo, o modelo fixo com estrutura da matriz de variâncias e covariâncias residuais ARH(1) e o modelo misto com estrutura da matriz de variâncias e covariâncias residuais ARH(1) . . . . .	52

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>13</b>
<b>2.1</b>	<b>PCR em tempo real</b>	<b>13</b>
<b>2.2</b>	<b>Tipos de quantificação</b>	<b>14</b>
<b>2.3</b>	<b>Análise dos dados de qPCR</b>	<b>15</b>
<b>2.3.1</b>	<b>Métodos tradicionais</b>	<b>17</b>
<b>2.3.1.1</b>	<b>Método <math>C_q</math> comparativo</b>	<b>18</b>
<b>2.3.1.2</b>	<b>Modelo de eficiência calibrada</b>	<b>19</b>
<b>2.3.2</b>	<b>Métodos estatísticos</b>	<b>20</b>
<b>2.3.2.1</b>	<b>Análise de variância considerando o modelo fixo</b>	<b>21</b>
<b>2.3.2.2</b>	<b>Modelos lineares mistos</b>	<b>24</b>
<b>2.3.2.3</b>	<b>Intervalo de confiança</b>	<b>28</b>
<b>2.3.3</b>	<b>Utilização dos métodos</b>	<b>30</b>
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>31</b>
<b>3.1</b>	<b>Material</b>	<b>31</b>
<b>3.2</b>	<b>Métodos</b>	<b>32</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>37</b>
<b>4.1</b>	<b>Modelo de eficiência calibrada</b>	<b>37</b>
<b>4.2</b>	<b>Método <math>C_q</math> comparativo</b>	<b>40</b>
<b>4.3</b>	<b>Análise de variância considerando o modelo fixo</b>	<b>43</b>
<b>4.4</b>	<b>Modelo misto</b>	<b>48</b>
<b>4.5</b>	<b>Comparação dos métodos aplicados</b>	<b>51</b>
<b>5</b>	<b>CONCLUSÕES</b>	<b>57</b>
	<b>REFERÊNCIAS</b>	<b>58</b>
	<b>APENDICE A – Rotina no programa R</b>	<b>62</b>
	<b>APENDICE B – Rotina no programa SAS</b>	<b>75</b>

## 1 INTRODUÇÃO

A reação em cadeia da polimerase (PCR) é uma técnica utilizada na amplificação de uma amostra de DNA *in vitro*, ou seja, sem a utilização de um organismo vivo. Uma extensão dessa técnica, denominada PCR em tempo real quantitativa (qPCR), permite que os processos de amplificação, detecção e quantificação do DNA produzidos sejam realizados em uma única etapa, o que faz com que ocorra maior agilidade na obtenção de resultados, diminuição no risco de contaminação da amostra e conseqüentemente uma maior precisão.

A qPCR tem sido amplamente utilizada na área médica e biológica para diagnóstico de doenças genéticas, detecção de agentes infecciosos, testes de paternidade, medicina forense, entre outros. A reação de transcrição reversa seguida da reação em cadeia da polimerase em tempo real quantitativa (RT-qPCR) é uma variação da técnica de qPCR e é empregada em experimentos de quantificação da expressão gênica.

Ao se realizar um experimento de qPCR, um valor denominado  $C_q$  (ciclo quantitativo) é determinado. O  $C_q$  representa o número de ciclos necessários para que o sinal de fluorescência atinja o limiar de detecção. A partir desse valor  $C_q$ , a quantificação absoluta e relativa dos ácidos nucleicos pode ser realizada. A quantificação relativa baseia-se na comparação da expressão de um gene alvo em relação a um gene de referência.

Vários procedimentos para a análise de dados para a quantificação relativa foram desenvolvidos. Entre eles, os mais utilizados são o método  $C_q$  comparativo e o modelo de eficiência calibrada. Por não envolverem métodos estatísticos, somente a utilização destes, pode fazer com que os pesquisadores obtenham inferências equivocadas acerca dos resultados obtidos. Portanto, os métodos tradicionais devem ser utilizados conjuntamente com métodos estatísticos para que se possa obter inferências válidas na análise de dados de qPCR.

A partir da necessidade da realização de uma análise estatística adequada, a qual possa ser utilizada para obter inferências válidas, diversos métodos estatísticos foram apresentados para serem utilizados na análise de dados de quantificação relativa.

O modelo de análise de variância com efeitos fixos e o modelo linear misto são métodos estatísticos que podem ser utilizados para a análise de dados de expressão gênica obtidos a partir da técnica de qPCR. No entanto, antes de se aplicar os métodos estatísticos citados, é primordial que se analise as pressuposições que cada método exige. Caso as pressuposições sejam violadas, as interpretações e inferências obtidas ficarão comprometidas.

Apesar de diversas metodologias terem sido desenvolvidas na análise de dados de qPCR, a escolha do melhor método a ser utilizado ainda é um assunto pouco discutido.

Este trabalho teve como objetivos aplicar e comparar métodos de análise de dados de expressão gênica obtidos a partir da técnica de qPCR, utilizando dados reais. Os métodos aplicados e posteriormente comparados foram:

- Método  $C_q$  comparativo;
- Modelo de eficiência calibrada;
- Análise de variância considerando o modelo fixo;
- Análise de variância considerando modelos lineares mistos.

## 2 REFERENCIAL TEÓRICO

### 2.1 PCR em tempo real

Quantificar os níveis de expressão gênica tornou-se um marco na maioria dos laboratórios de biologia molecular. Ao medir a quantidade de RNA celular, é possível determinar em que medida que o gene particular está sendo expresso. Para muitos genes, os níveis de expressão alteram drasticamente de gene para gene, célula para célula ou durante condições experimentais variadas (SCHMITTGEN; LIVAK, 2008).

A reação de transcrição reversa seguida da reação em cadeia da polimerase em tempo real quantitativa (RT-qPCR) é uma ferramenta poderosa para quantificar a expressão do gene (SCHMITTGEN; LIVAK, 2008; UDVARDI; CZECHOWSKI; SCHEIBLE, 2008). Segundo Nolan, Hands e Bustin (2006), a RT-qPCR é a combinação de três passos: (i) a conversão do RNA em DNA complementar (cDNA) por meio da transcrição reversa, (ii) a amplificação do cDNA utilizando a reação em cadeia da polimerase (PCR) e (iii) a quantificação da amplificação dos produtos de expressões de genes em tempo real.

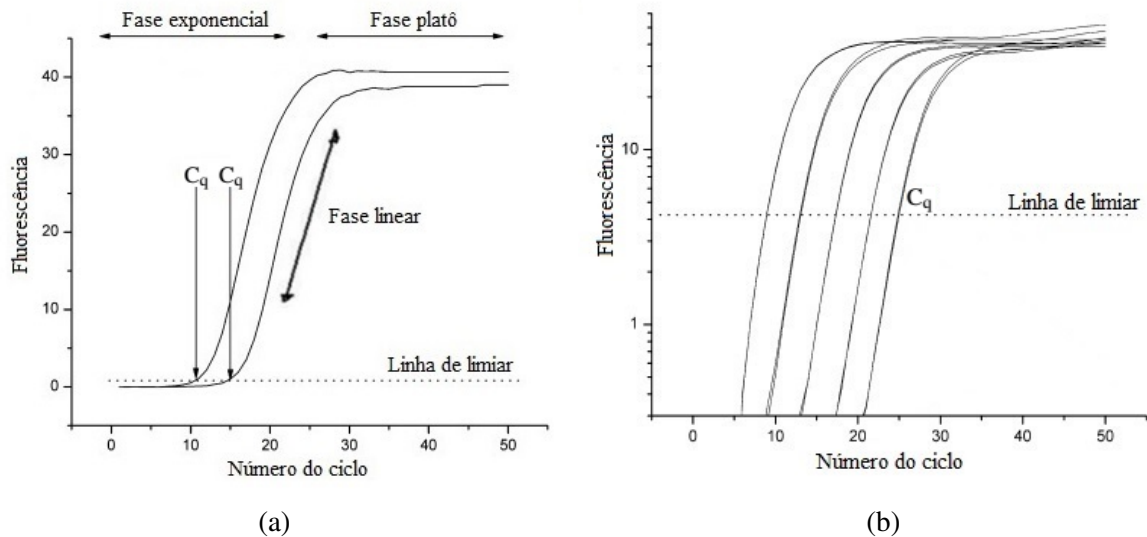
A PCR em tempo real quantitativa (qPCR) fornece a base para uma infinidade de aplicações em pesquisa básica, detecção de patógenos e diagnósticos biomédicos. Além disso, é amplamente aceita como padrão-ouro para a análise da expressão do gene (PABINGER et al., 2014). De acordo com Yuan et al. (2006), tanto DNA genômico quanto a transcrição reversa cDNA podem ser utilizados como moldes para a qPCR.

A técnica de qPCR é caracterizada por 3 principais fases: a fase exponencial, a fase linear e a fase platô. A dinâmica dessas fases consiste basicamente numa sucessão de ciclos de amplificação, no qual o ácido nucleico molde (*template*) é desnaturado, anelado com oligonucleotídeos iniciadores (*primers*) específicos e estendido para gerar uma cadeia complementar utilizando um DNA polimerase termoestável. Isso resulta em um aumento exponencial de amplicons (produtos de amplificação) que podem ser monitorados em cada ciclo (em tempo real), utilizando um repórter fluorescente. O aumento na fluorescência relacionado ao número do ciclo é plotado em um gráfico, para gerar a curva de amplificação, conforme pode ser observado na Figura 2.1(a) (POSTOLLEC et al., 2011). O gráfico, representado na Figura 2.1(b), refere-se ao sinal de fluorescência (transformado pelo logaritmo de base 2) em relação ao número de ciclos e apresenta uma extensão linear, na qual o sinal da fluorescência logaritmizado correlaciona-se com quantidade original de *templates*. Uma linha de limiar (*threshold line*) é definida e repre-



senta o limiar de detecção, ou seja, o número mínimo de ciclos para amplificação. O número do ciclo ao nível do limiar de fluorescência logaritmicado é denominado ciclo de quantificação  $C_q$  (muitas vezes denominado  $C_t$  para o ciclo limiar) (LOBO; LOBO, 2014; NASCIMENTO; SUAREZ; PINHAL, 2010; YUAN et al., 2006).

Figura 2.1 – Curvas de qPCR



Fonte: Adaptado de Kubista et al. (2006)

O  $C_q$  é o valor observado na maioria dos experimentos de qPCR e é a primeira medida de interesse estatístico (YUAN et al., 2006). O  $C_q$  sumariza os resultados e quanto maior o número de *template* no início da reação, um menor número de ciclos é necessário para alcançar um ponto, em que o sinal da fluorescência é primeiramente reconhecido como estatisticamente significativo, acima de um limiar arbitrário (*threshold*) (LOBO; LOBO, 2014).

Durante a fase exponencial o produto de PCR irá dobrar a cada ciclo, caso a eficiência for perfeita, ou seja, 100%. É possível obter uma eficiência de amplificação próxima a 100% na fase exponencial se as condições de PCR, as características do *primer*, a pureza do *template* e o comprimento dos amplicons forem ótimos (YUAN et al., 2006).

## 2.2 Tipos de quantificação

Utilizando o valor de  $C_q$ , a quantificação relativa e a quantificação absoluta podem ser realizadas. A quantificação relativa é baseada na expressão relativa de um gene alvo em relação a um gene de referência e é utilizada na maioria dos estudos de mudanças fisiológicas na

expressão de um gene (LIVAK; SCHMITTGEN, 2001; PFAFFL, 2001). A quantificação absoluta determina como, por exemplo, o número de cópias do transcrito de interesse, geralmente relacionando o sinal da PCR a uma curva padrão (LIVAK; SCHMITTGEN, 2001).

Ao se realizar a quantificação relativa, os genes de referência são comumente utilizados para normalizar a RT-qPCR, reduzindo possíveis erros gerados na quantificação da expressão de genes devido, por exemplo, à quantidade e à qualidade dos moldes de RNA e a eficiência da transcrição reversa (CHEN; FESSEHAIE; ARORA, 2012; PAOLACCI et al., 2009; VANDESOMPELE et al., 2002). Para um gene ser utilizado como referência ele deve apresentar expressão teoricamente constante, ou seja, sofrer pequena oscilação na expressão sob diferentes tratamentos (NASCIMENTO; SUAREZ; PINHAL, 2010).

Após a realização do tipo de quantificação, os tratamentos (geralmente diferentes grupos) são comparados estatisticamente para se avaliar a existência de possíveis diferenças significativas entre eles. Para possibilitar uma melhor análise estatística, gerando resultados significativos, Udvardi, Czechowski e Scheible (2008) recomendam a realização de três repetições independentes para cada tratamento e para cada amostra biológica são comumente utilizadas três réplicas, tanto para o gene alvo como para o gene referência (LOBO; LOBO, 2014). Caso o orçamento seja limitado para a realização de repetições e réplicas, as repetições devem ser priorizadas (PABINGER et al., 2014).

As réplicas (também denominadas repetições técnicas) são medições que são realizadas utilizando a mesma amostra para testar a reprodutibilidade da técnica de qPCR (instrumentos, reagentes ou protocolos). Ao se realizar a réplica, potenciais *outliers* são removidos. As repetições (também chamadas de repetições biológicas), por outro lado, são as medições extraídas em diferentes amostras de indivíduos de uma população, grupo ou tratamento, reproduzindo as condições experimentais e são realizadas para ser representante de uma condição biológica geral (GONI; GARCÍA; FOISSAC, 2009).

### **2.3 Análise dos dados de qPCR**

Desde que a quantificação relativa tornou-se objetivo da maioria dos experimentos que utilizam a qPCR, vários procedimentos têm sido sugeridos para a análise de dados de expressão gênica relativa (PFAFFL, 2001; SCHMITTGEN; LIVAK, 2008; STEIBEL et al., 2009; YUAN et al., 2006).

O modelo de eficiência calibrada proposto por Pfaffl (2001) e o método  $C_q$  comparativo (também conhecido com método  $\Delta\Delta C_q$  ou método  $2^{-\Delta\Delta C_q}$ ), proposto por Livak e Schmittgen (2001), são os mais utilizados para determinar a quantificação relativa de um gene alvo em comparação com um gene de referência (LOBO; LOBO, 2014; SCHMITTGEN; LIVAK, 2008). Esses métodos são baseados na razão da expressão relativa (*fold-change*), em que se comparam os níveis de expressão dos genes em dois grupos biológicos (tratado e controle por exemplo) (STEIBEL et al., 2009).

O modelo de eficiência calibrada e o método  $C_q$  comparativo, apesar de serem muito utilizados, não são métodos estatísticos em sentido estrito. Diversos trabalhos apresentam os resultados obtidos por estes métodos utilizando um gráfico de barras com indicação de um segmento de reta contendo o erro padrão ou o desvio padrão. No entanto, somente essas informações não podem ser utilizadas para se obter inferências válidas. Para se realizar uma análise adequada com a finalidade de se ter uma interpretação dos dados de qPCR mais robusta, considerações estatísticas são necessárias como testes de significância, intervalos de confiança e valor p (COFFEY; COFIEL, 2005; STEIBEL et al., 2009; YUAN et al., 2006). De acordo com Yuan et al. (2006) a não realização de uma modelagem e uma análise estatística adequada pode fazer com que os pesquisadores obtenham conclusões falso positivas na interpretação dos dados.

A necessidade de realização da análise estatística fez com que Yuan et al. (2006) apresentassem e comparassem em seu trabalho quatro metodologias para a análise de dados de qPCR, utilizando uma variação do método  $C_q$  comparativo. Os métodos utilizados foram a análise de variância, a análise de covariância, o teste t de Student e o teste de Wilcoxon-Mann-Whitney. Outro trabalho que utilizou metodologias estatísticas para análise foi o de Steibel et al. (2009). No artigo foi apresentado um método para a análise de dados de RT-qPCR no qual se utiliza um modelo linear misto e o comparou com o método  $C_q$  comparativo. Alguns outros trabalhos como os de Cui e Churchill (2003), Goni, García e Foissac (2009), Rieu e Powers (2009), Lobo e Lobo (2014) e Pabinger et al. (2014) também discutiram sobre métodos estatísticos para a análise de dados de expressão gênica.

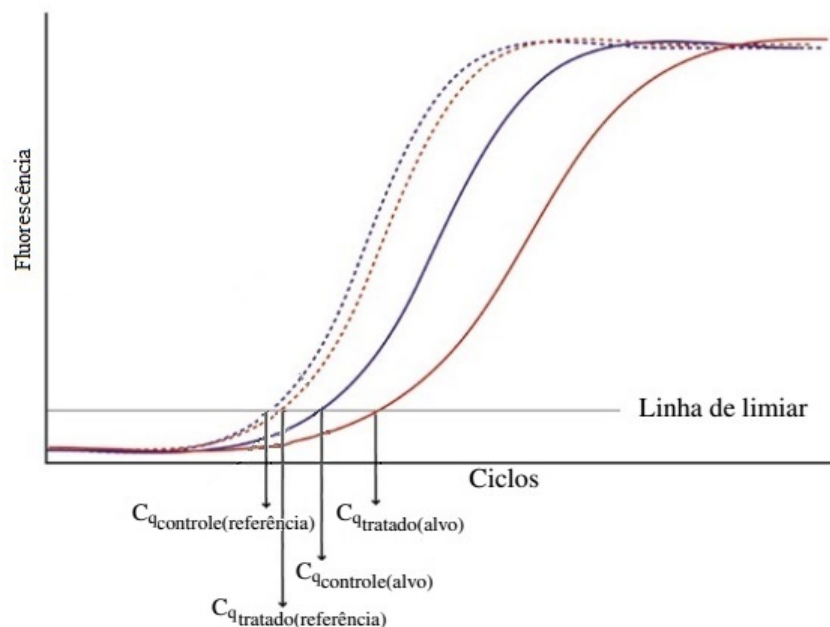
Apesar das metodologias citadas anteriormente serem utilizadas na análise dos dados de experimento de RT-qPCR, a escolha do melhor método para a análise ainda é um assunto pouco discutido. A seguir serão apresentadas algumas metodologias utilizadas na análise de expressão gênica.

### 2.3.1 Métodos tradicionais

Os métodos que são tradicionalmente utilizados na análise de dados de expressão gênica, obtidos por meio da técnica de qPCR são: o método  $C_q$  comparativo e o modelo de eficiência calibrada. De acordo com Yuan et al. (2006), o sistema experimental que envolve ambos os métodos é similar.

O experimento envolve uma amostra tratada e uma amostra controle e, para cada amostra, um gene alvo e um gene referência são incluídos para amplificação por PCR a partir de alíquotas diluídas em série (YUAN et al., 2006). Como pode ser observado na Figura 2.2, para cada amostra é obtido o valor de  $C_q$  para o gene alvo e para o gene referência que serão posteriormente comparados. Os diferentes  $C_q$ 's são definidos por:  $C_{q\text{controle(alvo)}}$  que representa o valor de  $C_q$  do gene alvo na amostra controle;  $C_{q\text{tratado(alvo)}}$  que representa o valor de  $C_q$  do gene alvo na amostra tratada;  $C_{q\text{controle(referência)}}$  que representa o valor de  $C_q$  do gene referência na amostra controle e  $C_{q\text{tratado(referência)}}$  que representa o valor de  $C_q$  do gene referência na amostra tratada.

Figura 2.2 – Quantificação por qPCR



Fonte: Adaptado de VanGuilder, Vrana e Freeman (2008)

Ao se realizar o experimento de qPCR normalmente várias repetições são incluídas para cada concentração diluída para se obter a eficiência de amplificação (YUAN et al., 2006).

A eficiência da amplificação ( $E$ ) é uma consideração importante quando se realiza a quantificação relativa, ela indica o desempenho geral do ensaio da qPCR (APPLIED BIOSYS-

TEMS, 2004; WONG; MEDRANO, 2005). Para obter o valor de E, calcula-se, após a realização do experimento de qPCR, os valores médios dos  $C_q$ 's obtidos a partir das repetições e então procede-se uma análise de regressão linear simples desses valores em função do logaritmo das concentrações do cDNA para a obtenção do *slope* (inclinação da reta) (LOBO; LOBO, 2014). O valor do *slope* é então aplicado na Equação 2.1 para se obter a eficiência de amplificação (PFAFFL, 2001).

$$E = 10^{\frac{-1}{slope}} \quad (2.1)$$

O valor de *slope* de -3.32 resulta em uma eficiência igual a 2, o que indica uma reação de PCR com 100% de eficiência, ou seja, a quantidade de produto formado deve dobrar a cada ciclo. Os *slopes* mais negativos que -3,32 indicam reações com menos de 100% de eficiência e *slopes* mais positivos que -3.32 podem indicar uma má qualidade da amostra ou problemas de pipetagem (APPLIED BIOSYSTEMS, 2004; VANGUILDER; VRANA; FREEMAN, 2008).

O método  $C_q$  comparativo e o modelo de eficiência calibrada, que serão apresentados a seguir, são diferentes apenas na maneira como consideram os valores de eficiência.

### 2.3.1.1 Método $C_q$ comparativo

No método  $C_q$  comparativo, presume-se que as eficiências de amplificação em tempo real dos genes alvo e referência sejam ótimas e idênticas, ou seja, sejam iguais a 2 (PFAFFL, 2001; YUAN et al., 2006).

Para obter a razão da expressão relativa para o método  $C_q$  comparativo primeiramente é necessário calcular para cada gene a diferença nos valores de  $C_q$  entre amostra tratada e controle e assim, obter o valor de  $\Delta C_q$  para o respectivo gene. Em seguida, a diferença entre os valores de  $\Delta C_q$  do gene alvo e  $\Delta C_q$  do gene referência é calculada, obtendo-se  $\Delta\Delta C_q$ . Este valor é então utilizado na Equação 2.2. A interpretação dessa razão pode ser dada como “a expressão do gene alvo relativo ao gene de referência na amostra tratada em comparação com a amostra controle” (SCHMITTGEN; LIVAK, 2008).

$$razão = 2^{-\Delta\Delta C_q}, \quad (2.2)$$

em que  $\Delta\Delta C_q = \Delta C_{q_{referência}} - \Delta C_{q_{alvo}}$  sendo  $\Delta C_{q_{alvo}} = C_{q_{controle(alvo)}} - C_{q_{tratado(alvo)}}$  e

$\Delta C_{q_{referência}} = C_{q_{controle(referência)}} - C_{q_{tratado(referência)}}$ .

Uma vantagem da utilização do método  $C_q$  comparativo é a facilidade de uso. A desvantagem é que deve-se assumir que a eficiência da PCR deve ser constante e ótima (SCHMITT-

GEN; LIVAK, 2008). De acordo com Lobo e Lobo (2014), normalmente aceita-se eficiência em torno de 90% a 110% ao se utilizar este método e para isso o *slope* obtido deve corresponder a um valor entre -3,6 e -3,1.

### 2.3.1.2 Modelo de eficiência calibrada

O modelo de eficiência calibrada combina a quantificação do gene e a normalização em um único cálculo. Esse modelo incorpora a eficiência de amplificação dos genes alvo e referência para corrigir as diferenças entre os dois ensaios (WONG; MEDRANO, 2005). A Equação 2.3 representa a razão da expressão relativa de um gene alvo comparado a um gene de referência, sendo  $E_{\text{alvo}}$  e  $E_{\text{referência}}$  as eficiências dos genes alvo e referência, respectivamente (YUAN et al., 2006; PFAFFL, 2001).

$$\text{razão} = \frac{(E_{\text{alvo}})^{\Delta C_{q_{\text{alvo}}}}}{(E_{\text{referência}})^{\Delta C_{q_{\text{referência}}}}}, \quad (2.3)$$

em que:  $\Delta C_{q_{\text{alvo}}} = C_{q_{\text{controle(alvo)}}} - C_{q_{\text{tratado(alvo)}}$  e  
 $\Delta C_{q_{\text{referência}}} = C_{q_{\text{controle(referência)}}} - C_{q_{\text{tratado(referência)}}$ .

Pode-se observar que o método  $C_q$  comparativo pode ser derivado do modelo de eficiência calibrada, caso as eficiências do gene alvo e do gene de referência forem ótimas e idênticas, ou seja,  $E_{\text{alvo}}=E_{\text{referência}}=2$  (YUAN et al., 2006). Dessa forma, a Equação 2.3 torna-se igual a Equação 2.2 (PFAFFL, 2001).

Uma das vantagens da utilização do modelo de eficiência calibrada, segundo Schmittgen e Livak (2008), é que a eficiência da PCR dos genes alvo e referência estão incluídos na Equação 2.3 e, portanto, diferenças na eficiência entre alvo e referência são contabilizadas no cálculo.

Como já citado anteriormente, uma outra desvantagem dos métodos que são tradicionalmente utilizados, como o método  $C_q$  comparativo e o modelo de eficiência calibrada, é que as interpretações obtidas somente por esses métodos, ou seja, sem considerações estatísticas, pode fazer com que os pesquisadores obtenham interpretações equivocadas acerca dos resultados obtidos nas análises de dados de RT-qPCR (COFFEY; COFIEL, 2005; STEIBEL et al., 2009; YUAN et al., 2006). Dessa maneira, devem ser utilizados juntamente com esses métodos, considerações estatísticas, tais como: testes de significância e intervalos de confiança (YUAN et al., 2006).

### 2.3.2 Métodos estatísticos

Como citado, os métodos tradicionais não utilizam de considerações estatísticas para se realizar inferências válidas na análise de dados de qPCR. Dessa maneira, para comparar dois grupos, alternativas estatísticas como o teste t de Student para médias de populações independentes e o teste de Wilcoxon-Mann-Whitney podem ser utilizados. Quando se deseja estudar dois ou mais grupos, um modelo linear pode ser utilizado para analisar os dados (PABINGER et al., 2014; STEIBEL et al., 2009).

Antes de realizar a análise dos dados de RT-qPCR é primordial que se satisfaça as pressuposições que o método exige. Pressuposições como a normalidade e homocedasticidade são necessárias em vários procedimentos estatísticos. Caso sejam violadas, as interpretações e inferências podem não ser válidas (LOBO; LOBO, 2014).

Diversos são os métodos utilizados para verificar a normalidade. Entre eles têm-se os métodos gráficos como Q-Q plot e histograma, que são mais subjetivos, pois dependem de interpretação visual, e os testes que apresentam um resultado mais objetivo como por exemplo: Qui-quadrado de Pearson, Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Shapiro-Francia, Cramer-von Mises e Anderson-Darling (TORMAN; COSTER; RIBOLDI, 2012).

Os testes utilizados para verificar a normalidade apresentam estatísticas de teste e metodologias diferentes, o que faz com que seja necessário um estudo mais aprofundado sobre qual é o mais adequado. Entretanto, eles têm em comum as seguintes hipóteses a serem testadas: a hipótese nula ( $H_0$ ) é de que a característica em estudo da população ou os erros seguem distribuição normal contra a hipótese alternativa ( $H_1$ ) de que a característica em estudo da população ou os erros não seguem distribuição normal (LOBO; LOBO, 2014; TORMAN; COSTER; RIBOLDI, 2012).

Para verificar a homogeneidade de variâncias os testes mais utilizados são os testes de Cochran, Levene, Hartley e Barlett. Nesses testes a hipótese nula de igualdade de variâncias é testada contra a hipótese alternativa de que nem todas as variâncias são iguais (ALMEIDA; ELIAN; NOBRE, 2008; LOBO; LOBO, 2014). Assim como os testes de normalidade, cada teste utilizado para a verificação da homocedasticidade possui metodologias e pressuposições diferentes, o que faz com que seja necessário um estudo mais aprofundado para a escolha do teste mais apropriado.

Caso haja a necessidade de analisar dados que não estejam de acordo com as pressuposições do modelo estatístico a ser utilizado, uma das alternativas a seguir deve ser utilizada:



modificar o método de análise, buscando aquele que melhor se ajusta aos dados, como um método não paramétrico; utilizar um modelo linear generalizado ou modificar a variável resposta realizando transformações, como por exemplo a transformação Box-Cox (BOX; COX, 1964), que será apresentada de forma mais detalhada a seguir (STORCK et al., 2000).

A transformação Box-Cox é uma família de transformações que pode ser realizada na variável resposta com a finalidade de se estabilizar ou reduzir a variabilidade existente e normalizar os resíduos (LÚCIO et al., 2011). Essa transformação está definida na Equação 2.4, em que  $y^{(\lambda)}$  é a variável transformada a partir de  $y$  para  $y > 0$ . Por meio da transformação Box-Cox espera-se identificar a melhor transformação na variável resposta, com base na utilização de um valor  $\lambda$  que maximize o estimador de máxima verossimilhança e minimize o resíduo (CHUNG; PEARN; YANG, 2007; LÚCIO et al., 2011).

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{para } \lambda \neq 0, \\ \log(y), & \text{para } \lambda = 0. \end{cases} \quad (2.4)$$

A seguir serão apresentados métodos estatísticos que podem ser utilizados para analisar dados de expressão gênica obtidos a partir da técnica de qPCR. Os métodos são: análise de variância considerando o modelo fixo, modelo linear misto e a utilização do intervalo de confiança.

### 2.3.2.1 Análise de variância considerando o modelo fixo

A análise de variância (ANOVA) é uma coleção de modelos estatísticos no qual a variância amostral é particionada em diversos componentes devido a diferentes fatores. Por meio dessa partição, a ANOVA estuda a influência desses fatores na característica de interesse. Dessa maneira, a ANOVA pode ser utilizada para comparar dois ou mais tratamentos e para verificar se determinado tratamento ou a interação entre dois ou mais fatores tem influência na expressão gênica (GONI; GARCÍA; FOISSAC, 2009). Cada um desses fatores presentes no modelo pode ser considerado como sendo de efeito fixo ou aleatório.

Define-se o fator de efeito fixo como uma variável categórica ou de classificação, para o qual o pesquisador inclui todos os níveis (ou condições) que são de interesse no estudo. Os níveis de um fator fixo são escolhidos de modo que eles representem condições específicas, e podem ser utilizados para definir contrastes (ou conjuntos de contrastes) de interesse no estudo de pesquisa (WEST; WELCH; GALECKI, 2014). O fator aleatório é uma variável de classifi-



cação, com níveis que representam uma amostra aleatória de uma grande população de níveis, podendo ser utilizado para fazer inferência sobre essa população (HEIBERGER; HOLLAND, 2004; SEARLE; CASELLA; MCCULLOCH, 2009; WEST; WELCH; GALECKI, 2014). Em contraste com os níveis de fatores fixos, os níveis de fatores aleatórios não representam condições escolhidas especificamente para atender aos objetivos do estudo (WEST; WELCH; GALECKI, 2014).

Os modelos de ANOVA em que há somente a presença de efeitos fixos são conhecidos como modelos de efeitos fixos. Quando há somente efeitos aleatórios, têm-se os modelos de efeitos aleatórios e quando há a presença de ambos os efeitos, os modelos são conhecidos como modelos mistos (SEARLE; CASELLA; MCCULLOCH, 2009).

Na ANOVA, os efeitos fixos presentes no modelo são testados por meio do teste F, que é um teste de razão de variâncias. Quando se aplica o teste F, uma das possíveis hipóteses que pode estar sendo testada é a hipótese nula ( $H_0$ ) de que não existe diferença significativa entre os níveis do fator. Quando o fator qualitativo apresentar somente dois níveis, o teste F significativo torna-se decisivo, pois ele indicará diferença significativa entre esses níveis, pois haverá somente um contraste a ser comparado. Quando o fator apresentar mais de dois níveis, os contrastes devem ser comparados por meio de testes de comparações múltiplas, como por exemplo, os testes de Tukey, Duncan, Scheffé, Dunnet e Bonferroni (BERTOLDO et al., 2007; LOBO; LOBO, 2014).

No caso em que o experimento seja conduzido em esquema fatorial, o teste F também pode ser utilizado para verificar se a interação entre os fatores é significativa. Caso seja significativa, há a indicação de que os fatores são dependentes. Assim, a análise dos efeitos principais não deve ser efetuada, devendo-se estudar o comportamento de um fator dentro de cada nível do outro fator, realizando o desdobramento dos graus de liberdade. Porém, se não for significativo, conclui-se que os fatores são independentes, ou seja, o comportamento de um fator independe da variação do outro fator e com isso as conclusões em separado são válidas (BERTOLDO et al., 2007; BARBIN, 2003; NOGUEIRA; CORRENTE, 2000; PERECIN; CARGNELUTTI FILHO, 2008).

Para exemplificar a utilização do modelo estatístico de ANOVA com efeitos fixos na análise de dados de RT-qPCR utiliza-se o exemplo apresentado no artigo de Yuan et al. (2006). A Equação 2.5 representa o modelo estatístico do experimento conduzido. Para este experimento foi utilizado o delineamento inteiramente casualizado, em esquema fatorial 2x2x4, sendo os fa-

tores representados por tratamento (tratado e controle), gene (alvo e referência) e concentração (0,08; 0,4; 2 e 10ng/ml). Foram realizadas três repetições. Então,

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}, \quad (2.5)$$

em que:

$y_{ijkl}$  representa o valor observado da variável resposta  $C_q$  no  $i$ -ésimo nível do fator tratamento, no  $j$ -ésimo nível do fator gene, na  $k$ -ésima concentração e na  $l$ -ésima repetição para  $i = 1, 2; j = 1, 2; k = 1, 2, 3, 4$  e  $l = 1, 2, 3$ ;

$\mu$  representa uma constante comum a todas as observações;

$\alpha_i$  representa o efeito do  $i$ -ésimo nível do fator tratamento;

$\beta_j$  representa o efeito do  $j$ -ésimo nível do fator gene;

$\gamma_k$  representa o efeito do  $k$ -ésimo nível do fator concentração;

$(\alpha\beta)_{ij}$  representa o efeito da interação entre  $i$ -ésimo nível do fator tratamento e o  $j$ -ésimo nível do fator gene;

$(\alpha\gamma)_{ik}$  representa o efeito da interação entre  $i$ -ésimo nível do fator tratamento e o  $k$ -ésimo nível do fator concentração;

$(\beta\gamma)_{jk}$  representa o efeito da interação entre  $j$ -ésimo nível do fator gene e o  $k$ -ésimo nível do fator concentração;

$(\alpha\beta\gamma)_{ijk}$  representa o efeito da interação tripla entre o  $i$ -ésimo nível do fator tratamento, o  $j$ -ésimo nível do fator gene e o  $k$ -ésimo nível do fator concentração;

$\varepsilon_{ijkl}$  representa o erro aleatório, com  $\varepsilon_{ijkl} \sim N(0, \sigma^2)$ .

De acordo com Yuan et al. (2006), ao se utilizar o modelo de ANOVA na análise de dados de RT-qPCR, o principal interesse é a interação gene e tratamento que irá fornecer a estimativa de  $\Delta\Delta C_q$ , ou seja,  $\Delta\Delta C_q$  poderá ser estimado a partir das diferentes combinações dos valores de  $(\alpha\beta)_{ij}$ . Dessa maneira, o objetivo é testar a hipótese nula de que a diferença no  $C_q$  entre os genes alvo e referência é a mesma nas amostras tratada e controle, isto é, deseja-se testar a seguinte hipótese:

$$H_0 : \mu_{CA} - \mu_{CR} = \mu_{TA} - \mu_{TR},$$

que é o mesmo que testar a  $H_0 : \mu_{CA} - \mu_{CR} - \mu_{TA} + \mu_{TR} = 0$  em que  $\mu_{CA}$ ,  $\mu_{CR}$ ,  $\mu_{TA}$  e  $\mu_{TR}$  é a média das combinações dos efeitos do gene alvo na amostra controle, do gene referência na amostra controle, do gene alvo na amostra tratada e do gene referência na amostra tratada,

respectivamente. Caso a hipótese nula não seja rejeitada, então  $\Delta\Delta C_q$  é significativamente igual a zero, e, conforme pode ser observado na Equação 2.2, se  $\Delta\Delta C_q$  for igual a zero, a razão terá como resultado 1, o que indica que não houve mudança na expressão do gene entre as amostras controle e tratada. Por outro lado, caso  $H_0$  seja rejeitada,  $\Delta\Delta C_q$  pode ser estimado a partir do contraste testado (YUAN et al., 2006). A partir do valor obtido para  $\Delta\Delta C_q$ , este pode então ser utilizado na Equação 2.2 para a obtenção da razão da expressão relativa.

Yuan et al. (2006) compararam o resultado obtido para a estimativa de  $\Delta\Delta C_q$  pelo método de ANOVA utilizando o modelo fixo com os resultados obtidos por outras alternativas existentes, como o teste t de Student e o teste não paramétrico de Wilcoxon-Mann-Whitney. Os autores obtiveram resultados semelhantes, mas eles afirmaram que a escolha do melhor método irá depender do delineamento experimental adotado, da qualidade do experimento e da verificação das pressuposições dos métodos.

Coffey e Cofiel (2005) também apresentaram comparações entre os testes que podem ser utilizados para análise da expressão gênica. Os autores citaram que o modelo de ANOVA fixo apresenta vantagens em relação aos testes t de Student e de Wilcoxon-Mann-Whitney. Como por exemplo, o modelo linear fixo permite a comparação de mais de dois grupos e o controle de outros efeitos fixos, que podem ser incluídos como termos adicionais no modelo, o que não são permitidos nos testes t de Student e de Wilcoxon-Mann-Whitney. Além disso, eles ainda complementam, que quando esses termos adicionais estão associados ao resultado, a variância residual pode ser substancialmente diminuída, o que permite maior poder na comparação dos dois grupos. E com isso, a contabilização de fontes conhecidas de variação aumenta o poder do experimento para observar diferenças significativas nos níveis de expressão para um determinado gene.

Para a utilização do modelo de ANOVA, as pressuposições de normalidade, homocedasticidade e independência dos erros e a aditividade dos termos do modelo devem ser satisfeitas.

### 2.3.2.2 Modelos lineares mistos

O modelo linear misto, apresentado na forma matricial na Equação 2.6, é um outro método para análise de dados que, de acordo com Steibel et al. (2009), pode ser utilizado na análise de expressão gênica envolvendo experimentos de RT-qPCR.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} , \quad (2.6)$$

em que:

$\mathbf{y}$  é um vetor  $n \times 1$  de  $n$  observações da variável resposta ( $C_q$ );

$\mathbf{X}$  é uma matriz  $n \times p$  de incidência dos  $p$  efeitos fixos (conhecida);

$\beta$  é um vetor  $p \times 1$  de efeitos fixos;

$\mathbf{Z}$  é uma matriz  $n \times q$  de incidência dos  $q$  efeitos aleatórios (conhecida);

$\mathbf{u}$  é um vetor  $q \times 1$  de efeitos aleatórios desconhecidos com  $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ ;

$\varepsilon$  é um vetor  $n \times 1$  de erros aleatórios com  $\varepsilon \sim N(\mathbf{0}, \mathbf{R})$ .

O modelo misto assume que os efeitos aleatórios e os erros seguem uma distribuição Normal com média zero e com matriz de variâncias e covariâncias dadas por  $\mathbf{G}$ , para os efeitos aleatórios e por  $\mathbf{R}$  para os erros. Assume-se também que  $\mathbf{u}$  e  $\varepsilon$  são não correlacionados (COFFEY; COFIEL, 2005).

As matrizes  $\mathbf{G}$  e  $\mathbf{R}$  podem assumir diferentes estruturas de variâncias e covariâncias como, por exemplo, as que consideram homogeneidade de variâncias, entre elas: autorregressiva de primeira ordem (AR(1)); simetria composta (CS); Toeplitz (TOEP) e a componentes de variância (VC). Outras que consideram a heterogeneidade de variâncias, são: autorregressiva heterogênea de primeira ordem (ARH(1)); simetria composta heterogênea (CSH); Toeplitz heterogênea (TOEPH) e a não estruturada (UN) (TORAL; ALENCAR; FREITAS, 2006). De acordo com Toral, Alencar e Freitas (2006), na estrutura CS todas as covariâncias são iguais e na VC todas são nulas. As demais estruturas consideram diferentes covariâncias para cada par de medidas.

Diversas técnicas podem ser utilizadas para a seleção do melhor modelo com a respectiva estrutura de covariância. Entre elas, de acordo com Littell et al. (2006), destacam-se os critérios de Akaike (Akaike's Information Criterion - AIC) (AKAIKE, 1974), o bayesiano de Schwarz (Bayesian Information Criterion- BIC) (SCHWARZ, 1978) e o teste de razão de verossimilhança.

Os modelos mistos são utilizados para descrever dados de experimentos cuja estrutura de tratamentos inclui os parâmetros de efeitos fixos e aleatórios. Os modelos em que há a presença de somente efeitos fixos, como o modelo apresentado por Yuan et al. (2006), são eficazes na análise de dados RT-qPCR. No entanto, eles assumem independência entre as observações e homogeneidade entre os níveis de expressão de um mesmo gene. Porém, ao se utilizar repetições e réplicas em um experimento de expressão gênica, espera-se que as observações realizadas numa mesma unidade sejam correlacionadas. No modelo fixo, para solucionar este problema

de correlação, é utilizado a média das réplicas e com essas médias realiza-se a análise. No entanto, essa abordagem faz com que não se utilize todas as informações disponíveis (CUI; CHURCHILL, 2003; COFFEY; COFIEL, 2005). De acordo com Cui e Churchill (2003), para esses casos, o modelo misto deve ser preferido. No modelo misto, as repetições devem ser tratadas como um fator aleatório e com isso podem ser obtidas inferências que são aplicadas às populações a partir da qual foram obtidas as amostras (CUI; CHURCHILL, 2003).

Além disso, é comum ao se realizar experimentos de expressão gênica, encontrar heterogeneidade de variâncias nas diversas ocasiões em que são feitas as medidas, o que faz com que a pressuposição de homocedasticidade do modelo linear clássico também não seja satisfeita. Dessa maneira, o modelo misto é uma opção que permite modelar tanto a correlação entre as medidas repetidas quanto a heterocedasticidade (BAUTISTA; QUARESMA; SANTANA, 2014; CUI; CHURCHILL, 2003; COFFEY; COFIEL, 2005).

Para exemplificar a utilização do modelo misto na análise de dados de RT-qPCR será considerado o trabalho de Steibel et al. (2009). O modelo representado na Equação 2.7 foi utilizado para estudar o efeito do gene inibidor da ligação do diazepam (DBI) no cérebro de leitões sujeito a dois tratamentos: desmame e isolamento social. O delineamento experimental utilizado foi em blocos casualizados, em esquema fatorial 2x2 sendo os fatores representados por desmame (amamentados e desmamados) e isolamento (isolado ou em ambiente controle) sendo utilizados animais provenientes de três leitegadas diferentes. O gene utilizado como referência foi o 18S e para todas as reações foram realizadas três repetições (referente ao animal, também denominado amostra no experimento), e para cada repetição foram realizadas três réplicas (referente ao poço da placa). O fator denominado tratamento consistiu na combinação dos fatores desmame e isolamento.

O modelo estatístico desse experimento é dado por:

$$y_{gijkr} = TG_{gi} + l_{gj} + B_{gijk} + D_{ijk} + \epsilon_{gijkr} , \quad (2.7)$$

em que:

$y_{gijkr}$  é o valor observado da variável resposta  $C_q$  do  $g$ -ésimo gene sujeito ao  $i$ -ésimo tratamento, da  $j$ -ésima leitegada, da  $r$ -ésima réplica referente a  $k$ -ésima repetição para  $g= 1, 2$ ;  $i= 1, 2, 3, 4$ ;  $j= 1, 2, 3$ ;  $k= 1, 2, 3$  e  $r= 1, 2, 3$ .

$TG_{gi}$  representa o efeito do  $i$ -ésimo nível do fator tratamento na expressão do  $g$ -ésimo gene;

$l_{gj}$  representa o efeito aleatório do  $g$ -ésimo gene na  $j$ -ésima leitegada com  $l_{gj} \sim N(0, \sigma_{l_g}^2)$ ;  
 $B_{gijk}$  representa o efeito aleatório da  $k$ -ésima repetição dentro do  $g$ -ésimo gene, do  $i$ -ésimo tratamento e da  $j$ -ésima leitegada com  $B_{gijk} \sim N(0, \sigma_{B_g}^2)$ ;

$D_{ijk}$  representa o efeito da  $k$ -ésima repetição dentro do  $i$ -ésimo nível do fator tratamento na  $j$ -ésima leitegada com  $D_{ijk} \sim N(0, \sigma_D^2)$ ;

$\varepsilon_{gijk}$  representa o erro aleatório, com  $\varepsilon_{gijk} \sim N(0, \sigma^2)$ .

Para o modelo apresentado na Equação 2.7, o objetivo, assim como na ANOVA com modelo fixo, é testar a hipótese nula de que a diferença no  $C_q$  entre os genes alvo e referência é a mesma nas amostras tratada e controle. Portanto, a interação gene e tratamento é o efeito de interesse.

Steibel et al. (2009) compararam o modelo apresentado na Equação 2.7 com o método  $C_q$  comparativo. Os autores concluíram que o modelo proposto por eles apresenta como vantagem em relação ao método  $C_q$  comparativo, a possibilidade de testar qualquer hipótese linear geral, como por exemplo, a possibilidade de se testar a hipótese de interação entre o isolamento social e o desmame precoce na expressão do DBI no cérebro de leitões. Com o método  $C_q$  comparativo não se tem a possibilidade de testar a mesma hipótese, pois sua aplicação é limitada a comparação de pares de tratamentos.

Os autores também compararam o modelo misto proposto com os métodos propostos por Yuan et al. (2006). Eles observaram que os métodos propostos por Yuan et al. (2006) foram utilizados para análise com apenas um nível de repetição, sendo que na prática muitos experimentos de RT-qPCR incluem tanto as repetições quanto as réplicas. Assim como foi observado por Cui e Churchill (2003) e Coffey e Cofiel (2005), Steibel et al. (2009) discutiram que quando há a presença de repetições e réplicas, utilizar as médias das réplicas, como proposto por Yuan et al. (2006), não é a melhor opção. Para esses casos, o modelo misto pode ter um melhor desempenho.

Outro exemplo de trabalho que utilizou a proposta do Steibel et al. (2009) para o ajuste de modelo misto nos dados de RT-qPCR foi o de Busato et al. (2016). Neste, os autores queriam avaliar a expressão de 13 genes relacionados na regulação do *turnover* proteico no músculo *longissimus dorsi* de bovinos das raças Angus e Nelore. Além destes trabalhos, o modelo misto é também utilizado para a análise de dados de RT-qPCR em Benítez et al. (2015), González-Calvo et al. (2014), McCulloch et al. (2014) e Lau e Lennon (2011).

### 2.3.2.3 Intervalo de confiança

De acordo com Tichopad et al. (2013), a maneira mais comum de visualização dos resultados da análise de dados de expressão gênica obtidas por meio da técnica de qPCR, é a utilização de um diagrama de barras com segmentos de retas associados a essas barras. A utilização do intervalo de confiança, como segmento de reta, é uma maneira de ilustrar a precisão da média estimada e principalmente, indicar a significância estatística de forma gráfica (TICHOPAD et al., 2013).

O intervalo de confiança é uma amplitude de valores, derivados de estatística de amostra, no qual julga-se com um risco conhecido de erro, conter o valor de um parâmetro populacional. A seguir será apresentado uma definição mais formal para a obtenção do intervalo de confiança.

Seja  $\theta$  um parâmetro populacional desconhecido. Pode-se definir a partir de uma amostra  $\mathbf{x}=(x_1, \dots, x_n)$  intervalos de confiança para  $\theta$  dado um nível de confiança  $100(1-\alpha)\%$  sendo  $0 \leq \alpha \leq 1$ . Estes intervalos podem ser escritos como  $[L(\mathbf{x}); U(\mathbf{x})]$  em que  $L(\mathbf{x})$  e  $U(\mathbf{x})$  são estatísticas e representam respectivamente os limites inferiores e superiores do intervalo de confiança (CASELLA; BERGER, 2002; SMITHSON, 2003). Dessa maneira, tem-se que

$$P[L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})] = 1 - \alpha,$$

ou seja, o parâmetro desconhecido  $\theta$  apresenta uma probabilidade de  $1-\alpha$  de estar entre os limites  $L(\mathbf{x})$  e  $U(\mathbf{x})$  (FERREIRA, 2009). Segundo Andrade (2011), se o intervalo  $[L(\mathbf{x}); U(\mathbf{x})]$  é pequeno, tem-se um elevado grau de precisão da inferência realizada.

De acordo com Ferreira (2009), o método da função pivô é o mais comum para a obtenção de regras gerais para a construção dos limites, inferior e superior, de um intervalo de confiança. A função pivô depende de  $\theta$  e de seu estimador  $\hat{\theta}$ , mas sua distribuição de probabilidade não depende de  $\theta$  (FERREIRA, 2009).

No caso em que o interesse é obter o intervalo de confiança para a média  $\mu$  de uma população através de uma amostra, considera-se dois casos distintos: quando a variância populacional  $\sigma^2$  é conhecida e quando esta é desconhecida. A seguir serão apresentados estes dois casos.

Supondo que deseja-se obter o intervalo de confiança para  $\mu$  de uma população com distribuição normal e  $\sigma^2$  conhecida. Tem-se que a distribuição de  $\bar{X}$  é uma distribuição normal

com média  $\mu$  e variância  $\frac{\sigma^2}{n}$ , ou seja,  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ . Dessa maneira tem-se que:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}},$$

em que  $Z$  é a função pivô, que tem distribuição normal padrão, isto é,  $Z \sim N(0, 1)$ . A distribuição de  $Z$  não depende de  $\mu$ , logo é possível obter os quantis inferior e superior  $\frac{\alpha}{2}$ , dados respectivamente por  $-Z_{\frac{\alpha}{2}}$  e  $Z_{\frac{\alpha}{2}}$  (FERREIRA, 2009). Então, tem-se que

$$P\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Substituindo a definição de  $Z$  na expressão anterior e isolando  $\mu$ , obtém-se

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha. \quad (2.8)$$

Segundo Ferreira (2009), a afirmativa probabilística 2.8 significa que o parâmetro desconhecido  $\mu$  apresenta uma probabilidade  $1 - \alpha$  de estar entre os limites  $\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$  e  $\bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ . O intervalo 2.8 é utilizado quando o valor de  $\sigma^2$  é conhecido. No entanto, na maioria das situações reais, o valor da variância populacional não é conhecido e o intervalo 2.8 tem-se sua utilidade prática limitada (FERREIRA, 2009). Para os casos em que a variância populacional é desconhecida utiliza-se a variância amostral  $S^2$  no lugar de  $\sigma^2$ . A função pivô é dada por

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}},$$

que tem distribuição t de Student com  $n - 1$  graus de liberdade. Então, de maneira semelhante à obtida para o caso de  $\sigma^2$  conhecida, é apresentado na afirmativa probabilística 2.9, o intervalo com  $100(1 - \alpha)\%$  para o parâmetro desconhecido  $\mu$  para o caso em que a variância  $\sigma^2$  é desconhecida. Assim,

$$P\left(\bar{X} - t_{((n-1), \frac{\alpha}{2})} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{((n-1), \frac{\alpha}{2})} \frac{S}{\sqrt{n}}\right) = 1 - \alpha. \quad (2.9)$$

Logo,

$$\text{IC}(\mu, 1 - \alpha) = \left[ \bar{X} - t_{((n-1), \frac{\alpha}{2})} \frac{S}{\sqrt{n}}; \bar{X} + t_{((n-1), \frac{\alpha}{2})} \frac{S}{\sqrt{n}} \right]. \quad (2.10)$$



De acordo com Steibel et al. (2009) utilizando as informações apresentadas na Equação 2.10, é possível construir intervalos de confiança a  $100(1-\alpha)\%$  para o verdadeiro valor de  $\Delta\Delta C_q$  e para a razão da expressão relativa obtidas pelo modelo de eficiência calibrada e pelo método  $C_q$  comparativo.

Ao serem utilizados modelos de ANOVA para a análise de dados de RT-qPCR, González-Calvo et al. (2014) e Benítez et al. (2015) apresentaram em seus trabalhos uma alternativa para a obtenção do intervalo de confiança para a razão da expressão relativa. Estes autores utilizaram a aproximação Normal para a obtenção do intervalo de confiança a 95% para o verdadeiro valor da razão. Os limites inferior e superior do intervalo de confiança são obtidos utilizando a seguinte expressão:

$$IC = 2^{-(\Delta\Delta C_q \pm 1,96 \times EP)}, \quad (2.11)$$

em que EP refere-se ao erro padrão da estimativa de  $\Delta\Delta C_q$ .

### 2.3.3 Utilização dos métodos

De acordo com uma pesquisa realizada no Portal de Periódicos CAPES, foram encontrados 35.736 artigos relacionados a RT-qPCR, o que demonstra que a técnica tem sido bastante utilizada por diversos pesquisadores. Ao se realizar a pesquisa neste mesmo portal, foi observado após redirecionamento para a página do PubMed, que o artigo de Livak e Schmittgen (2001), referente a utilização do método  $C_q$  comparativo, e do Pfaffl (2001), referente ao modelo de eficiência calibrada, foram citados em 21077 e 6257 artigos, respectivamente. Dessa maneira, verifica-se que o método  $C_q$  comparativo tem sido mais utilizado que o modelo de eficiência calibrada. No Portal de Periódicos CAPES foram observadas a utilização do modelo misto na análise de dados de experimentos que envolvem a RT-qPCR em 248 artigos o que demonstra que a utilização deste tipo de método ainda não está bem difundida entre os pesquisadores.

### 3 MATERIAL E MÉTODOS

#### 3.1 Material

O conjunto de dados de expressão gênica obtido pela técnica de qPCR que será utilizado nas análises refere-se a um experimento realizado por Lima (2015), no qual foram utilizadas plantas de café (*Coffea arabica*) da cultivar ‘Catuaí Vermelho’.

O experimento foi conduzido em uma estufa semi-controlada na Universidade Federal de Lavras, em Lavras, Minas Gerais, e teve como objetivo avaliar as mudanças na regulação do etileno do cafeeiro, pela análise da expressão de genes de sua biossíntese e sinalização, em curtos intervalos de tempo, após reidratação ou aplicação do produto químico 1-MCP, que atua como um inibidor da ação do etileno. As plantas de café (*Coffea arabica*) da cultivar ‘Catuaí Vermelho’ utilizadas foram cedidas com 6 meses, pela Fundação Procafé, localizada em Varginha, Minas Gerais.

A RT-qPCR foi utilizada para analisar os níveis de expressão de três genes da biossíntese e de um gene do receptor do etileno do cafeeiro, denominados, respectivamente, por CaACS1-like, CaACO1-like, CaACO4-like e CaETR4-like. Foram considerados cinco diferentes tratamentos: plantas bem irrigadas (WW), plantas com déficit de água (WD), plantas reidratadas (RW), plantas tratadas com 1-MCP (1-MCP) e plantas controle com 1-MCP (BTH). O tratamento BTH consistiu na aplicação do produto químico *Break-Thru* juntamente com o 1-MCP. O *Break-Thru* faz com que ocorra um aumento da penetração do 1-MCP.

As plantas que receberam os tratamentos WW, 1-MCP e BTH foram mantidas bem irrigadas, recebendo irrigação regular. As plantas que receberam os tratamentos WD e RW tiveram a irrigação suspensa até que o potencial hídrico foliar chegasse a -2,0Mpa.

As análises de expressão dos prováveis genes relacionados às rotas de biossíntese e sinalização do etileno do cafeeiro foram realizadas em folhas e raízes, em quatro tempos diferentes (0, 2, 6 e 24 horas após a instituição dos tratamentos). No tempo 0, a análise fisiológica e a amostragem do material foram realizadas somente nas plantas que receberam os tratamentos WW e WD. Os tratamentos RW, 1-MCP e BTH foram aplicados no tempo 0. Os níveis de expressão gênica foram analisados nas folhas considerando os tratamentos WW, WD, RW, 1-MCP e BTH e nas raízes somente com os tratamentos WW, WD e RW.

Para as análises foram realizadas três repetições (também denominadas repetições biológicas) de cada tratamento, em cada um dos tempos, e para cada repetição foram realizadas

três réplicas (também denominadas repetições técnicas). A Ubiquitina foi utilizada como gene de referência.

Para ilustrar a utilização dos métodos de análise de dados de RT-qPCR foi selecionado para esse trabalho o gene CaACS1-like analisado na raiz.

### 3.2 Métodos

Os métodos tradicionais juntamente com os métodos estatísticos para análise de dados de expressão gênica relativa foram utilizados para analisar os dados obtidos por Lima (2015), referentes ao gene CaACS1-like, analisado na raiz. Para a aplicação dos métodos foram desconsideradas das análises o tempo 0, pois somente os tratamentos WW e WD haviam sido observados nesse tempo e para que as técnicas estatísticas pudessem ser aplicadas e comparadas, todos os tratamentos deveriam ser analisados em tempos iguais. A amostra calibradora definida para as análises dos genes foi o tratamento WW no tempo 2.

Inicialmente foram analisados os dados de expressão gênica pelo modelo de eficiência calibrada e pelo método  $C_q$  comparativo utilizando o *software* R (R Core Team, 2016). Para ambos os métodos,  $C_q$  comparativo e modelo de eficiência calibrada, foram obtidos os valores de  $\Delta C_q$  para os respectivos genes alvo e referência, subtraindo-se os valores médios de  $C_q$  das réplicas da amostra tratada do valor médio de  $C_q$  referente as repetições da amostra considerada como calibradora (ou controle), como realizado por Lima (2015).

Para a análise utilizando do método  $C_q$  comparativo, foram obtidos os valores de  $\Delta\Delta C_q$  a partir do valor médio das diferenças entre o  $\Delta C_q$  do gene alvo e o  $\Delta C_q$  do gene referência para cada tratamento presente na análise. Utilizando essas diferenças foram obtidos os erros padrões ( $EP_{\Delta\Delta C_q}$ ) e os intervalos de confiança ( $IC_{\Delta\Delta C_q}$ ) a 95% para  $\Delta\Delta C_q$ , como proposto por Steibel et al. (2009). A significância das estimativas de  $\Delta\Delta C_q$  obtidas pelo método  $C_q$  comparativo foram determinadas utilizando o teste t de Student. Para construir o intervalo de confiança a 95% para  $\Delta\Delta C_q$ , o qual pode ser utilizado para verificar a diferença entre o gene alvo e referência em cada tratamento, foi utilizada a seguinte expressão:

$$IC_{\Delta\Delta C_q} = \Delta\Delta C_q \pm 4,3026 \times EP_{\Delta\Delta C_q},$$

em que 4,3026 refere-se ao o quantil 97,5 da distribuição t de Student com 2 graus de liberdade (referentes ao número de repetições menos 1).

Uma vez determinadas as diferenças entre o  $\Delta C_q$  do gene alvo e o  $\Delta C_q$  do gene referência e com isso obtido os valores de  $\Delta\Delta C_q$ , esses valores foram aplicados na fórmula  $2^{-\Delta\Delta C_q}$  e, a partir do valor médio, foi obtido os valores das razões da expressão gênica relativa. Foram obtidos também os erros padrões ( $EP_{Razão}$ ) e o intervalo de confiança ( $IC_{Razão}$ ) a 95% para essas razões. Para a obtenção dos intervalos de confiança para essas razões, que podem ser utilizados para verificar as diferenças na expressão do gene CaACS1-like entre os tratamentos aplicados, foi utilizada a seguinte expressão:

$$IC_{Razão} = Razão \pm 4,3026 \times EP_{Razão}. \quad (3.1)$$

Para a utilização do método  $C_q$  comparativo, conforme especificado por Livak e Schmittgen (2001), considerou-se a eficiência dos genes alvo e referência como sendo igual a dois.

Para a análise utilizando o modelo de eficiência calibrada, foram consideradas as eficiências dos genes alvo e referência determinadas por Lima (2015). Para esse método, os valores obtidos para  $\Delta C_q$  e eficiências para o genes alvo e referência foram utilizados na fórmula  $\frac{(E_{alvo})^{\Delta C_{qalvo}}}{(E_{referência})^{\Delta C_{qreferência}}}$ , conforme proposto por Pfaffl (2001). A partir do valor médio foi obtido o valor da razão da expressão gênica relativa. Foram obtidos os erros padrões e os intervalos de confiança de 95% para essas razões. Os intervalos de confiança para as razões calculadas pelo modelo de eficiência calibrada foram obtidos assim como para o método  $C_q$  comparativo, ou seja, utilizando a expressão 3.1.

Para ambos os métodos, modelo de eficiência calibrada e  $C_q$  comparativo, ao se realizar o intervalos de confiança para  $\Delta\Delta C_q$  e razão, primeiramente foram realizados testes de Shapiro-Wilk para verificar a normalidade dos dados.

Após a realização das análises pelo método  $C_q$  comparativo e pelo modelo de eficiência calibrada, foram realizadas as análises utilizando o modelo fixo de análise de variância e, em seguida, utilizando modelos lineares mistos. Para a análise por ambos os métodos (modelos fixos e modelos mistos) foi considerado o fator denominado trat, que consistiu na combinação fatorial (3x3) do tratamento (RW, WD e WW) e do tempo (2, 6 e 24).

Para o modelo de ANOVA com a presença de somente efeitos fixos foram ajustados dois tipos de modelos. Inicialmente as análises foram realizadas considerando como valores de  $C_q$  a média das réplicas em cada repetição ( $C_q$  médio). Em seguida, ajustou-se um segundo modelo que considerou os valores de  $C_q$  referentes a todas as réplicas.

Para o modelo ajustado com o  $C_q$  médio, utilizando o *software* R, foram realizadas as verificações das pressuposições a 5% de significância, por meio de testes apropriados, como o de Shapiro-Wilk para a verificação da normalidade e o de Bartlett para a avaliação da homogeneidade de variâncias. Para que a pressuposição de homogeneidade de variâncias pudesse ser atendida, foi realizada a transformação Box-Cox na variável resposta  $C_q$  médio utilizando a função *boxcox* do pacote *MASS*, do *software* R. Para o modelo ajustado utilizando todos os valores de  $C_q$  como havia a presença de heterogeneidade de variâncias dos erros e a presença de medidas repetidas em uma mesma planta, foram utilizadas diferentes matrizes de variâncias e covariâncias residuais para que se pudesse modelar a heterogeneidade de variâncias existente e a correlação entre as medidas repetidas. Para a seleção do modelo com a respectiva estrutura de covariância que melhor se ajustou aos dados foram utilizados os critérios de informação de Akaike (AIC) e de Schwarz (BIC). O modelo escolhido, com a respectiva estrutura de covariância, foi aquele que apresentou menor valor de AIC e de BIC.

O modelo fixo de ANOVA ajustado considerando o  $C_q$  médio foi o seguinte:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

em que:

$y_{ijk}$  representa o valor observado da variável resposta  $C_q$  médio no  $i$ -ésimo nível do fator trat, no  $j$ -ésimo nível do fator gene, na  $k$ -ésima repetição para  $i = 1, \dots, 9$ ;  $j = 1, 2$  e  $k = 1, 2, 3$ .

$\mu$  representa uma constante comum a todas as observações;

$\alpha_i$  representa o efeito do  $i$ -ésimo nível do fator trat;

$\beta_j$  representa o efeito do  $j$ -ésimo nível do fator gene;

$(\alpha\beta)_{ij}$  representa o efeito da interação entre o  $i$ -ésimo nível do fator trat e o  $j$ -ésimo nível do fator gene;

$\varepsilon_{ijk}$  representa o erro aleatório e independente com  $\varepsilon_{ijk} \sim N(0, \sigma^2)$ . Assim, o vetor de erros  $\varepsilon \sim N_n(\mathbf{0}, \mathbf{I}\sigma^2)$ .

O modelo fixo de ANOVA ajustado considerando todos os valores de  $C_q$  foi o seguinte:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijkl},$$

em que:

$y_{ijkl}$  representa o valor observado da variável resposta  $C_q$  no  $i$ -ésimo nível do fator trat, no  $j$ -ésimo nível do fator gene, na  $l$ -ésima réplica da  $k$ -ésima repetição para  $i = 1, \dots, 9$ ;  $j = 1, 2$ ;  $k = 1, 2, 3$  e  $l = 1, 2, 3$ .

$\mu$  representa uma constante comum a todas as observações;

$\alpha_i$  representa o efeito do  $i$ -ésimo nível do fator trat;

$\beta_j$  representa o efeito do  $j$ -ésimo nível do fator gene;

$(\alpha\beta)_{ij}$  representa o efeito da interação entre o  $i$ -ésimo nível do fator trat e o  $j$ -ésimo nível do fator gene;

$\varepsilon_{ijkl}$  representa o erro aleatório, sendo que o vetor de erros  $\varepsilon \sim N_n(\mathbf{0}, \mathbf{R})$ .

Ao se ajustar o modelo misto, foram consideradas as plantas (repetições) como efeito aleatório, como sugerido por Steibel et al. (2009). Portanto, o modelo misto ajustado foi o seguinte:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_{k(ij)} + \varepsilon_{ijkl},$$

em que:

$y_{ijkl}$  representa o valor observado da variável resposta  $C_q$  no  $i$ -ésimo nível do fator trat, no  $j$ -ésimo nível do fator gene, na  $l$ -ésima réplica da  $k$ -ésima repetição para  $i = 1, \dots, 3$ ,  $j = 1, 2$ ,  $k = 1, 2, 3$  e  $l = 1, 2, 3$ .

$\mu$  representa uma constante comum a todas as observações;

$\alpha_i$  representa o efeito do  $i$ -ésimo nível do fator trat;

$\beta_j$  representa o efeito do  $j$ -ésimo nível do fator gene;

$(\alpha\beta)_{ij}$  representa o efeito da interação entre o  $i$ -ésimo nível do fator trat e o  $j$ -ésimo nível do fator gene;

$\delta_{k(ij)}$  representa o efeito aleatório da  $k$ -ésima planta dentro do  $j$ -ésimo nível do fator gene e do  $i$ -ésimo nível do fator trat, com  $\delta_{k(ij)} \sim N(0, \sigma_\delta^2)$ ;

$\varepsilon_{ijkl}$  representa o erro aleatório, sendo que o vetor de erros  $\varepsilon \sim N_n(\mathbf{0}, \mathbf{R})$ .

Da mesma maneira como realizado para o modelo fixo, que considera todos os valores de  $C_q$  presentes no conjunto de dados, foram consideradas para o modelo misto, diferentes matrizes de variâncias e covariâncias residuais para que se pudesse modelar a heterogeneidade de variâncias e a correlação entre as medidas repetidas presentes no experimento. Para a seleção do modelo com a respectiva estrutura de covariância que melhor se ajustou aos dados foram utilizados os critérios de informação de Akaike (AIC) e de Schwarz (BIC).

Ao se ajustar os modelos fixos e misto, para aqueles que apresentaram o efeito da interação  $(\alpha\beta)_{ij}$  significativo a 5%, foram realizados, como sugerido por Yuan et al. (2006) e Steibel et al. (2009), diferentes contrastes de interesse entre os níveis de trat e gene para se obter as estimativas de  $\Delta\Delta C_q$ . Para cada estimativa de  $\Delta\Delta C_q$  foram obtidos o erro padrão e o intervalo de confiança assimétrico (95%) como apresentado por Steibel et al. (2009). A significância das estimativas de  $\Delta\Delta C_q$  foram determinadas utilizando o teste t de Student.

Os valores estimados de  $\Delta\Delta C_q$  foram utilizados na fórmula  $2^{-\Delta\Delta C_q}$ , como sugerido por Steibel et al. (2009), para se obter as estimativas das razões. Para cada estimativa da razão obtida, foram calculados intervalos de confiança a 95%. Estes intervalos foram obtidos utilizando uma modificação do intervalo de confiança apresentado por González-Calvo et al. (2014) e Benítez et al. (2015) mostrado na Equação 2.11. Para a obtenção do intervalo de confiança para as razões, os autores consideraram a aproximação Normal. Para este trabalho foi considerada para a construção dos intervalos a distribuição t de Student, como sugerido por Steibel et al. (2009). Para o modelo misto, os limites superior e inferior dos inferior dos intervalos foram obtidos utilizando a seguinte expressão:

$$IC = 2^{-(\Delta\Delta C_q \pm 1,98 \times EP)}. \quad (3.2)$$

em que 1,98 refere-se ao quantil 97,5 da distribuição t de Student com 139 graus de liberdade e EP refere-se ao erro padrão da estimativa de  $\Delta\Delta C_q$ .

Para o modelo fixo foi obtido o intervalo de confiança a 95% para cada razão utilizando a seguinte expressão:

$$IC = 2^{-(\Delta\Delta C_q \pm 2,13 \times EP)}. \quad (3.3)$$

em que 2,13 refere-se ao quantil 97,5 da distribuição t de Student com 15 graus de liberdade e EP refere-se ao erro padrão da estimativa de  $\Delta\Delta C_q$ .

As análises para os modelos fixos e misto ajustados foram realizadas utilizando o software R e o *proc mixed* do programa estatístico SAS.

## 4 RESULTADOS E DISCUSSÃO

No presente trabalho serão apresentados os resultados obtidos por quatro diferentes métodos de análise de dados de PCR em tempo real: modelo de eficiência calibrada; método  $C_q$  comparativo; análise de variância considerando o modelo fixo e a análise de variância considerando o modelo misto.

Ao se utilizar o modelo de eficiência calibrada serão consideradas no cálculo as eficiências de amplificação dos genes alvo e referência determinadas por Lima (2015). Para o método  $C_q$  comparativo serão consideradas as eficiências dos genes alvo e referência como sendo ótimas e idênticas, ou seja, iguais a 2. Este método foi considerado aceitável de ser utilizado uma vez que as eficiências de ambos os genes estavam entre 0,9 e 1,1, intervalo de aceitação para a utilização do método  $C_q$  comparativo. Para as análises pelos modelos de ANOVA, fixo e misto, serão consideradas as eficiências de amplificação como ótimas, visto que para o cálculo da razão é utilizada a fórmula do método  $C_q$  comparativo.

Após a realização das análises, será realizada uma comparação dos resultados obtidos referentes às razões de expressão relativa, afim de verificar o método que apresentou melhores resultados, tendo como base a amplitude do intervalo de confiança obtido para a razão.

### 4.1 Modelo de eficiência calibrada

Os resultados das análises de expressão gênica realizadas pelo modelo de eficiência calibrada são apresentados na Tabela 4.1 e na Figura 4.1. As eficiências determinadas por Lima (2015) para os genes Ubiquitina e CaACS1-like foram 1,99 e 1,90, respectivamente.

Na Tabela 4.1 têm-se as estimativas das razões para o gene CaACS1-like em relação ao gene Ubiquitina, observadas nos tempos 2, 6 e 24h, quando aplicados os tratamentos bem irrigado (WW), déficit de água (WD) e reirrigado (RW). Para cada uma dessas estimativas são apresentados o erro padrão, o valor-p do teste de normalidade de Shapiro-Wilk e o intervalo com 95% de confiança para as médias das razões.

O valor-p do teste de normalidade de Shapiro-Wilk é utilizado para verificar se a amostra usada para a obtenção do intervalo de confiança provém de uma distribuição normal. Quando o valor-p for menor que o nível de significância adotado, rejeita-se a hipótese de normalidade dos dados e o intervalo de confiança não é válido, pois a normalidade é uma pressuposição para a obtenção do intervalo de confiança. Apesar do teste de Shapiro-Wilk poder ser utilizado para a verificação da normalidade em pequenas amostras, este necessita de no mínimo três



amostras para sua aplicação. O tratamento WD no tempo 2, continha apenas duas repetições, logo o valor-p para essa análise não pode ser obtido, o que faz com que o intervalo de confiança também não possa ser calculado. Para as demais combinações de tratamento e tempo presentes na Tabela 4.1, verifica-se que a hipótese de normalidade não foi rejeitada a 5% de significância, com isso, os intervalos de confiança para as razões são válidos e podem ser utilizados.

Tabela 4.1 – Estimativas das razões da expressão relativa do gene CaACS1-like obtidas pelo modelo de eficiência calibrada para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada razão são apresentados o erro padrão, o valor-p do teste de Shapiro-Wilk e o intervalo de confiança a 95%

Tratamento	Tempo	Razão	Erro padrão	Valor-p	Intervalo de confiança
RW	2	0,8206	0,2278	0,85244	<b>(-0,1595; 1,8008)</b>
WD	2	0,4020	0,0018	†	††
WW	2	1,0031	0,0553	0,73962	<b>(0,7653; 1,2409)</b>
RW	6	0,9030	0,0573	0,73523	<b>(0,6564; 1,1495)</b>
WD	6	1,1103	0,3650	0,48976	<b>(-0,4602; 2,6807)</b>
WW	6	0,9983	0,2016	0,13754	<b>(0,1309; 1,8657)</b>
RW	24	0,5173	0,1415	0,20312	<b>(-0,0915; 1,1262)</b>
WD	24	0,4887	0,0662	0,46692	<i>(0,2039; 0,7736)</i>
WW	24	0,6135	0,1416	0,08897	<b>(0,0043; 1,2228)</b>

†Teste de Shapiro Wilk não aplicado. ††Intervalo de confiança não obtido

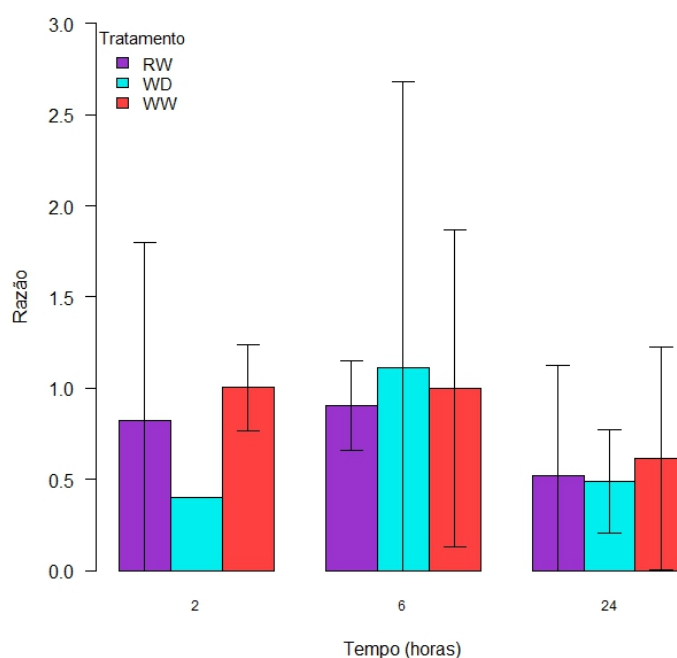
Os intervalos de confiança para a razão podem ser utilizados para verificar se existe diferença significativa entre os genes alvo e referência em cada tratamento. Os intervalos que não contêm o valor 1 indicam existir diferença significativa entre estes genes. Na Tabela 4.1 pode ser observado que os genes CaACS1-like e Ubiquitina são diferentemente expressos na combinação de tratamento e tempo em que o intervalo de confiança é apresentado em itálico. Os intervalos de confiança que contém o valor 1, apresentados na Tabela 4.1 em negrito, indicam que para a combinação tratamento e tempo analisada, a expressão dos genes não apresentou diferença. Cabe ressaltar que alguns intervalos de confiança obtidos possuem grande amplitude, que ocorreram devido a alta variabilidade existente entre as repetições (média das réplicas) e quanto maior a amplitude menor a precisão nas estimações das médias. Uma das possíveis soluções para aumentar a precisão na estimação das médias, seria realizar um maior número de repetições.

Além de serem utilizado para a verificação da diferença significativa entre os genes alvo e referência em cada tratamento, os intervalos de confiança podem ser utilizados para se obter inferências válidas na análise de dados de RT-qPCR, ao comparar a expressão do gene alvo entre os tratamentos. Caso os intervalos de confiança para a razão se sobreponham, há

evidências de que a expressão do gene alvo normalizado para o gene de referência não difere entre os tratamentos comparados. Alguns trabalhos que utilizam os métodos tradicionais para a análise de dados de expressão gênica, como é o caso de Faibish, Suzuki e Bartlett (2016), Pradeepa et al. (2016) e Huang, Xu e Cheng (2014), utilizam somente a sobreposição dos erros padrões para se obter as interpretações da significância dos resultados obtidos. No entanto, o erro padrão fornece uma estimativa da precisão da média e somente a sua utilização juntamente com a razão não pode ser empregada, pois não é uma forma correta de inferência. É necessário o emprego de métodos estatísticos adequados para realizar inferências, como é o caso do intervalo de confiança.

A Figura 4.1 representa nas barras os valores referentes as razões (*Fold-change*) do gene CaACS1-like analisados na raiz, obtidas pelo modelo de eficiência calibrada, sendo que os segmentos de retas sobrepostos às barras representam os intervalos de confiança.

Figura 4.1 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas pelo modelo de eficiência calibrada para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões



Ao se analisar a Figura 4.1, observa-se que quando ocorre sobreposição dos segmentos referentes aos intervalos de confiança a expressão do gene alvo normalizado para o gene referência não difere entre os tratamentos comparados. Para o tempo 2, verifica-se na Figura 4.1, que o intervalo de confiança para a razão referente ao tratamento WD não é apresentado,

pois como já especificado, não foi possível verificar a normalidade dos dados, necessária para a obtenção do intervalo de confiança. Portanto, a comparação da expressão do gene CaACS1-like entre os tratamentos WW e WD para este tempo não pode ser realizada. Ao se comparar a expressão do gene CaACS1-like no tempo 2 ao se utilizar os tratamentos RW e WD, nota-se que esta não diferiu. Esta análise pode ser realizada pois o intervalo de confiança obtido para a razão referente ao tratamento RW se sobrepôs a barra referente a razão do tratamento WD. Para as análises ocorridas nos tempos 6 e 24, observa-se que a expressão do gene CaACS1-like não diferenciou entre os tratamentos aplicados. Vale salientar que, para a análises realizadas, observa-se que os intervalos de confiança obtidos possuem grande amplitude o que faz com que a precisão dos resultados obtidos seja baixa. Para que se possa obter maior precisão nas conclusões sobre a significância dos resultados obtidos utilizando o modelo de eficiência calibrada, um maior número de repetições é necessário.

#### **4.2 Método $C_q$ comparativo**

As análises de expressão gênica realizadas pelo método  $C_q$  comparativo são apresentadas nas Tabelas 4.2 e 4.3 e na Figura 4.2. Nas Tabelas 4.2 e 4.3 estão representadas, respectivamente, as estimativas de  $\Delta\Delta C_q$  e razão obtidas por meio do método  $C_q$  comparativo para o gene CaACS1-like. Para cada uma dessas estimativas são apresentados o erro padrão, o valor-p do teste de normalidade de Shapiro-Wilk e o intervalo com 95% de confiança para as médias de  $\Delta\Delta C_q$  e razões. Para a estimativa de  $\Delta\Delta C_q$  é apresentado também na Tabela 4.2 o valor-p do teste t de Student. O teste t de Student é utilizado para testar a hipótese nula de que a diferença no  $C_q$  entre os genes alvo e referência é a mesma nas amostras tratada e controle, ou seja, testar hipótese de que  $\Delta\Delta C_q$  é igual a zero.

Utilizando o valor-p do teste de Shapiro-Wilk, observa-se nas Tabelas 4.2 e 4.3, que a normalidade dos dados referentes a  $\Delta\Delta C_q$  e a razão foi rejeitada a 5% de significância para o tratamento WW no tempo 24. Não foi possível, para o tratamento WD no tempo 2, determinar os valores-p do teste de Shapiro-Wilk referentes a  $\Delta\Delta C_q$  e razão devido ao baixo número de repetições disponíveis, pois como já especificado, o teste de Shapiro-Wilk exige um número mínimo de 3 amostras para sua aplicação. Assim, para os valores em que não pode ser obtido o valor-p ou que a hipótese de normalidade é rejeitada, os intervalos de confiança e o teste t-de Student não são válidos, pois para a utilização do intervalo de confiança ou aplicação do teste t de Student, a amostra utilizada deve ser obtida de uma população com distribuição normal. Para

os demais valores presentes nas Tabelas 4.2 e 4.3 verifica-se que a hipótese de normalidade não foi rejeitada e portanto, os intervalos de confiança para  $\Delta\Delta C_q$  e razão e os valores-p dos testes t de Student são válidos e podem ser utilizados.

Tabela 4.2 – Estimativas de  $\Delta\Delta C_q$  obtidas pelo método  $C_q$  comparativo para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa de  $\Delta\Delta C_q$  são apresentados erro padrão, o valor-p do teste de Shapiro-Wilk, o valor-p do teste t de Student e o intervalo de confiança a 95%

Tratamento	Tempo	$\Delta\Delta C_q$	Erro padrão	Valor-p <sup>1</sup>	Valor-p <sup>2</sup>	Intervalo de confiança
RW	2	0,4478	0,4268	0,90301	0,36325	<b>(-1,3886; 2,2842)</b>
WD	2	1,4528	0,0041	†	††	††
WW	2	0	0,0928	0,66682	1	<b>(-0,3993; 0,3993)</b>
RW	6	0,1989	0,0897	0,6911	0,19821	<b>(-0,187; 0,5848)</b>
WD	6	0,1444	0,6667	0,33244	0,84933	<b>(-2,7242; 3,013)</b>
WW	6	0,0911	0,3281	0,1795	0,80253	<b>(-1,3206; 1,5028)</b>
RW	24	1,1256	0,3755	0,3097	0,05367	<b>(-0,49; 2,7412)</b>
WD	24	1,1567	0,2162	0,49863	0,00795	<i>(0,2265; 2,0869)</i>
WW	24	0,8833	0,4339	0,02965	††	††

<sup>1</sup>Valor-p teste de Shapiro-Wilk. <sup>2</sup>Valor-p do teste t de Student. †Teste de Shapiro-Wilk não aplicado. ††Teste t de Student não aplicado e intervalo de confiança não obtido.

Pode-se verificar nas Tabelas 4.2 e 4.3, que alguns intervalos de confiança possuem grandes amplitudes, o que também foi observado ao se realizar as análises pelo modelo de eficiência calibrada. Como, por exemplo, o intervalo obtido para a razão, observado na Tabela 4.3, quando a análise do gene CaACS1-like ocorreu no tempo 6 ao ser utilizado o tratamento WD. As grandes amplitudes ocorreram devido a alta variabilidade existente entre as repetições e como já citado, uma das possíveis soluções para diminuir essa amplitude seria realizar um maior número de repetições ao se fazer o experimento.

Tabela 4.3 – Estimativas das razões da expressão relativa do gene CaACS1-like obtidas pelo método  $C_q$  comparativo para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa da razão são apresentados o erro padrão, o valor-p do teste de Shapiro-Wilk e o intervalo de confiança a 95%

Tratamento	Tempo	Razão	Erro padrão	Valor-p	Intervalo de confiança
RW	2	0,7978	0,2238	0,81947	<b>(-0,1651; 1,7607)</b>
WD	2	0,3653	0,001	†	††
WW	2	1,0041	0,0635	0,72084	<b>(0,7309; 1,2773)</b>
RW	6	0,8746	0,0551	0,6394	<b>(0,6375; 1,1117)</b>
WD	6	1,0813	0,3765	0,5903	<b>(-0,5386; 2,7012)</b>
WW	6	0,9916	0,243	0,12569	<b>(-0,0539; 2,0371)</b>
RW	24	0,4923	0,1377	0,20722	<b>(-0,1002; 1,0848)</b>
WD	24	0,4583	0,0645	0,60454	<i>(0,1808; 0,7358)</i>
WW	24	0,5873	0,1451	0,045	††

†Teste de Shapiro Wilk não aplicado. ††Intervalo de confiança não obtido

Ao se utilizar o método  $C_q$  comparativo pode-se utilizar três métodos diferentes para verificar se existe diferença significativa entre os genes alvo e referência em cada tratamento. Pode-se utilizar o intervalo de confiança obtido para  $\Delta\Delta C_q$  ou o intervalo de confiança obtido para razão ou o valor-p do teste t de Student. Na prática deve-se optar por uma das maneiras apresentadas anteriormente. Caso seja utilizado o intervalo de confiança para  $\Delta\Delta C_q$ , deve-se verificar se o valor 0 encontra-se incluído no intervalo de confiança, caso esteja incluído há evidências de que a expressão dos genes não difere para o tratamento analisado. Quando utilizado o intervalo de confiança para a razão, deve-se observar se o valor 1 encontra-se presente no intervalo, se o valor 1 está incluído, há evidências de que a expressão dos genes alvo e referência seja a mesma. Caso o método escolhido for analisar utilizando o valor-p do teste t de Student deve-se verificar se o valor-p é menor que o nível de significância ( $\alpha$ ) adotado. Caso o valor-p seja menor que  $\alpha$  há evidências de diferença significativa na expressão dos genes alvo e referência. Dessa maneira, pode ser observado na Tabela 4.2 ou na Tabela 4.3 que a expressão dos genes CaACS1-like e Ubiquitina é similar para as combinações de tratamento e tempo em que o intervalo de confiança é apresentado em negrito. Para o intervalo apresentado em ambas as tabelas em itálico, há evidências de que os genes são diferentemente expressos.

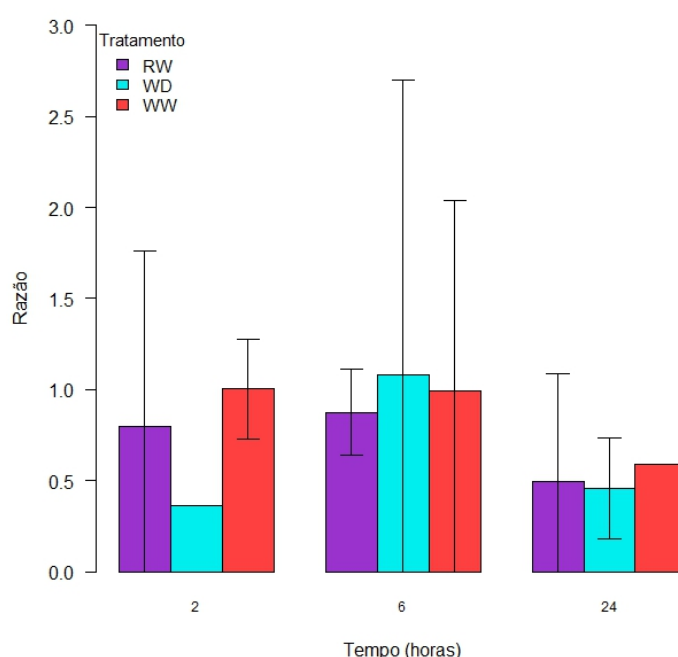
Como já citado, alguns trabalhos que realizam as análises de dados de expressão gênica pelos métodos tradicionais, usam de forma equivocada a sobreposição dos erros padrões para se obter conclusões nas análises de dados de RT-qPCR. Dessa forma, os pesquisadores podem obter conclusões erradas a respeito da significância dos resultados obtidos.

Na Figura 4.2 representam-se nas barras os valores referentes as razões (*Fold-change*) do gene CaACS1-like analisados na raiz obtidas pelo método  $C_q$  Comparativo, sendo que os segmentos de retas sobrepostos às barras representam os intervalos de confiança.

Observa-se na Figura 4.2 que quando ocorre sobreposição dos segmentos referentes aos intervalos de confiança, a expressão do gene alvo não difere entre os tratamentos comparados. Para o tempo 2, não é apresentado o intervalo de confiança para a razão ao ser utilizado o tratamento WD, pois não foi possível verificar a normalidade dos dados, necessária para a obtenção do intervalo de confiança. Para o tratamento WW analisado no tempo 24 o intervalo de confiança não é apresentado, pois como já especificado, a hipótese de normalidade foi rejeitada e portanto o intervalo de confiança não pode ser obtido. Dessa maneira, para o tempo 2, a comparação da expressão do gene CaACS1-like entre os tratamentos WW e WD não pode ser realizada. Para as demais combinações aos pares de tratamentos observa-se que para os tem-

pos 2, 6 e 24 não houve diferença na expressão do gene CaACS1-like. Assim como observado ao serem realizadas as análises pelo modelo de eficiência calibrada, os intervalos de confiança obtidos apresentaram grandes amplitudes e para se ter maior precisão nas análises, um maior número de repetições deve ser realizado.

Figura 4.2 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas pelo método  $C_q$  comparativo para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões



### 4.3 Análise de variância considerando o modelo fixo

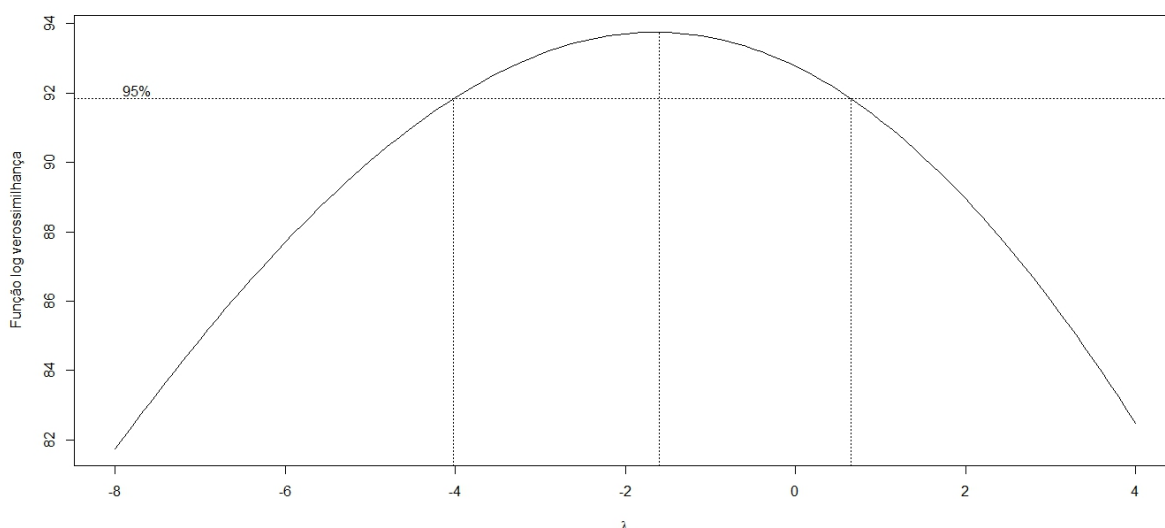
Para a análise de variância considerando o modelo fixo, inicialmente foram considerados como variável resposta o valor de  $C_q$  referente à média das réplicas ( $C_q$  médio) e como fatores os efeitos fixos de gene, trat e da interação trat\*gene. Para este modelo, foram consideradas as médias das réplicas para que a premissa de independência entre as observações pudesse ser atendida, visto que as réplicas são medidas realizadas em uma mesma unidade amostral, no caso, em uma mesma planta do café.

Para o modelo ajustado utilizando o  $C_q$  médio, foram verificadas as pressuposições de normalidade e homogeneidade de variâncias dos resíduos. Ao nível de 5% de significância o teste de Shapiro-Wilk indicou que os resíduos apresentavam distribuição normal (valor-p = 0,2933) e pelo teste de Bartlett foi observado que estes eram heterocedásticos (valor-p =

0,0348). Como foi verificada a violação da pressuposição de homogeneidade de variâncias dos resíduos, as interpretações e inferências obtidas pelo modelo ajustado podem não ser válidas. Para que a pressuposição de homogeneidade de variâncias pudesse ser atendida, foi realizada a transformação Box-Cox (BOX; COX, 1964) na variável resposta  $C_q$  médio.

A transformação Box-Cox consiste em encontrar um valor de  $\lambda$  que define o tipo de transformação mais adequada para que as pressuposições de normalidade e homogeneidade possam ser atendidas (BOX; COX, 1964). No *software* R para se obter o valor de  $\lambda$  e, assim, realizar a transformação, foi utilizado a função *boxcox* do pacote *MASS*, que forneceu o gráfico apresentado na Figura 4.3.

Figura 4.3 – Estimação do parâmetro  $\lambda$  pelo método de Box-Cox para o modelo fixo de ANOVA utilizando o  $C_q$  médio



Ao observar a Figura 4.3, verifica-se que o valor de  $\lambda$  que maximiza a função de verossimilhança é aproximadamente -1,7. No entanto, observa-se que o valor 0 encontra-se incluído no intervalo de confiança de 95%, o que sugere que a transformação logarítmica pode ser utilizada. Dessa maneira, ajustou-se novamente o modelo utilizando como variável resposta o logaritmo de  $C_q$  médio.

Após a transformação foram verificadas novamente as pressuposições. Ao nível de 5% de significância pelo teste de Shapiro-Wilk foi observado que os resíduos continuaram sendo normalmente distribuídos (valor-p = 0,2665) e pelo teste de Bartlett que estes passaram a apresentar homogeneidade de variâncias (valor-p = 0,0544). Na Tabela 4.4 é apresentado o resumo da análise de variância para o modelo ajustado após a transformação Box-Cox.

Tabela 4.4 – Resumo da análise de variância para o modelo fixo utilizando como valor de  $C_q$  a média das réplicas após transformação Box-Cox

Fonte de variação	Grau de liberdade	Valor-p
Trat	8	0,0643
Gene	1	<0,0001
Trat*Gene	8	0,2383

Pode-se observar pela Tabela 4.4 a evidência de significância para o efeito de gene. Para a interação trat\*gene não foi detectado efeito significativo. Dessa maneira, há evidências de que os fatores são independentes, ou seja, o comportamento dos genes independe da variação das combinações dos tratamentos e tempos aplicados. Portanto, como somente o efeito de gene foi significativo, verifica-se que para todas as combinações de tratamento e tempo analisados o gene CaACS1-like obteve maior valor médio de  $C_q$  (23,4146) em relação à Ubiquitina (20,2326). Porém, como neste caso o interesse é comparar a expressão do gene alvo entre os tratamentos, a interação não significativa indica que para todas as combinações de tratamento e tempo analisados a expressão do gene CaACS1-like foi a mesma. Dessa maneira, constata-se que a expressão do gene CaACS1-like não diferiu entre os tratamentos ao se ajustar o modelo fixo utilizando o  $C_q$  médio.

Após a realização das análises utilizando o valor de  $C_q$  médio e verificado que este modelo não apresentou efeito significativo da interação trat\*gene, ajustou-se um novo modelo de efeitos fixos utilizando como valor de  $C_q$  todas as réplicas. Para o modelo agora ajustado foi observada, pelo teste de Bartlett, a heterogeneidade de variâncias dos resíduos (valor-p<0,0001). Além disso, devido a presença de todas as réplicas, verifica-se a presença de dependência entre as medidas realizadas em uma mesma planta o que prejudica a análise estatística. Para contornar estes problemas, foram consideradas diferentes estruturas da matriz de variâncias e covariâncias residuais. O modelo com a matriz de variâncias e covariâncias residuais sob estrutura do tipo autorregressiva heterogênea de primeira ordem (ARH(1)) foi o que apresentou melhor ajuste aos dados, obtendo menor valor para AIC e para BIC.

A escolha da matriz de variâncias e covariâncias residuais ARH(1) é coerente para esta análise, pois esta matriz modela a heterogeneidade de variâncias de todas as combinações de trat e gene e a dependência dos erros, visto que esta considera a desigualdade de variâncias e covariâncias e reconhece a existência de correlação entre as medidas repetidas. O resumo dos resultados da análise de variância para o modelo ajustado sob estrutura ARH(1) da matriz de variâncias e covariâncias residuais, é apresentado na Tabela 4.5.



Tabela 4.5 – Resumo da análise de variância para o modelo fixo com a estrutura de matriz de variâncias e covariâncias residuais do tipo ARH(1) para o conjunto de dados utilizando como valor de  $C_q$  todas as réplicas

Fonte de Variação	Grau de liberdade	Valor-p
Trat	8	<0,0001
Gene	1	0,0005
Trat*Gene	8	<0,0001

Observa-se na Tabela 4.5 que a interação trat\*gene apresentou efeito significativo, o que evidencia que os fatores atuam de forma dependente, ou seja, o comportamento dos genes depende da variação das combinações dos tratamentos e tempos aplicados. Como a interação apresentou efeito significativo, foram realizados contrastes, como sugerido por Yuan et al. (2006) e Steibel et al. (2009), para se obter as estimativas de  $\Delta\Delta C_q$ . As estimativas de  $\Delta\Delta C_q$  obtidas para cada combinação de tratamento e tempo aplicados são apresentados na Tabela 4.6. Para cada uma dessas estimativas são apresentados o erro padrão, o valor-p do teste t de Student para a hipótese de que  $\Delta\Delta C_q$  é igual a 0 e o intervalo de confiança a 95%.

Tabela 4.6 – Estimativas de  $\Delta\Delta C_q$  obtidas pelo modelo fixo com estrutura de matriz de variâncias e covariâncias residuais ARH(1) ajustado utilizando o  $C_q$  referente a todas as réplicas, para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa de  $\Delta\Delta C_q$  são apresentados o erro padrão, o valor-p do teste t de Student e o intervalo de confiança a 95%

Tratamento	Tempo	$\Delta\Delta C_q$	Erro padrão	Valor-p	Intervalo de confiança
RW	2	0,3414	0,2783	0,2388	<b>(-0,2518; 0,9347)</b>
WD	2	1,7923	0,2403	<0,0001	(1,2801; 2,3045)
WW	2	†	†	†	†
RW	6	0,1788	0,09688	0,0848	<b>(-0,02772; 0,3853)</b>
WD	6	0,1508	0,3368	0,6607	<b>(-0,567; 0,8686)</b>
WW	6	0,03696	0,2785	0,8962	<b>(-0,5566; 0,6305)</b>
RW	24	1,1103	0,1779	<0,0001	(0,7312; 1,4894)
WD	24	1,0703	0,139	<0,0001	(0,7741; 1,3665)
WW	24	0,7571	0,2359	0,0058	(0,2544; 1,2598)

† Amostra calibradora

Assim como especificado no método  $C_q$  comparativo, o valor-p do teste t de Student ou o intervalo de confiança para  $\Delta\Delta C_q$  apresentados na Tabela 4.6 podem ser utilizados para verificar se a expressão dos genes alvo e referência diferem em cada combinação de tratamento e tempo presente na análise. Caso seja utilizado o intervalo de confiança para a  $\Delta\Delta C_q$  a expressão dos genes não é diferente caso o valor 0 esteja presente no intervalo de confiança. Caso utiliza-se o valor-p do teste t de Student, a expressão dos genes é diferente quando o valor-p é menor

que o nível de significância ( $\alpha$ ) adotado. Dessa maneira, observa-se na Tabela 4.6 que para as combinação de tratamento e tempo em que os intervalos de confiança são apresentados em itálico, há evidências de que houve diferença significativa na expressão dos genes CaACS1-like e Ubiquitina. Para as combinação de tratamento e tempo em que os intervalos de confiança são apresentados em negrito, há evidências de que os genes se expressaram de forma similar.

Utilizando os valores estimados de  $\Delta\Delta C_q$  apresentados na Tabela 4.6, foram obtidas as razões da expressão relativa por meio da fórmula  $2^{-\Delta\Delta C_q}$ . As razões, com os respectivos intervalos de confiança obtidos conforme apresentado na Equação 3.3, são apresentados na Tabela 4.7 e na Figura 4.4.

Tabela 4.7 – Estimativas das razões da expressão relativa do gene CaACS1-like utilizando os valores de  $\Delta\Delta C_q$  estimados pelo modelo fixo com estrutura de matriz de variâncias e covariâncias residuais ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa da razão são apresentados o intervalo de confiança a 95%

<b>Tratamento</b>	<b>Tempo</b>	<b>Razão</b>	<b>Intervalo de confiança</b>
RW	2	0,7893	<b>(0,5232; 1,1907)</b>
WD	2	0,2887	<i>(0,2024; 0,4118)</i>
WW	2	†	†
RW	6	0,8834	<b>(0,7656; 1,0194)</b>
WD	6	0,9008	<b>(0,5477; 1,4815)</b>
WW	6	0,9747	<b>(0,6459; 1,4708)</b>
RW	24	0,4632	<i>(0,3561; 0,6024)</i>
WD	24	0,4762	<i>(0,3878; 0,5848)</i>
WW	24	0,5917	<i>(0,4176; 0,8384)</i>

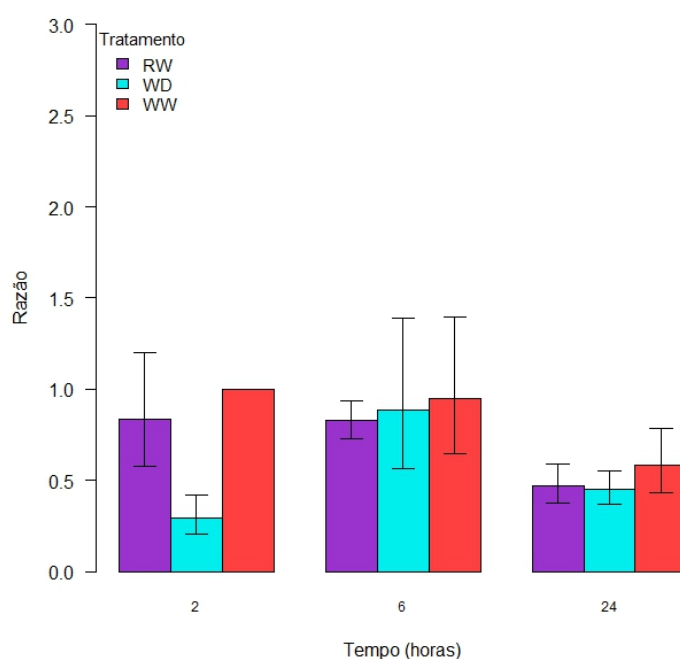
†Amostra calibradora

Da mesma maneira como utilizado para os métodos  $C_q$  comparativo e modelo de eficiência calibrada, os intervalos de confiança para a razão apresentados na Tabela 4.7 também podem ser utilizados para verificar se existe diferença significativa na expressão dos genes CaACS1-like e Ubiquitina em cada tratamento. Para as combinações tratamento e tempo em que os intervalos de confiança contem o valor 1, apresentados na Tabela 4.7 em negrito, há a indicação que a expressão dos genes é similar. Para os intervalos de confiança que não contém o valor 1, apresentados em itálico, há evidências de que a expressão dos genes é diferente. Vale salientar novamente que, apesar de existir três métodos diferentes para se verificar a existência de diferença significativa entre os genes alvo e referência em cada tratamento, deve-se escolher somente um método para realizar a análise.

Os intervalos de confiança para a razão também podem ser utilizados para comparar a expressão do gene alvo entre os tratamentos. Como pode ser observado na Figura 4.4, para

os intervalos de confiança que ocorre a sobreposição, há evidências de que os tratamentos não diferem em relação a expressão do gene CaACS1-like. Para o tempo 2, verifica-se na Figura 4.4, que o intervalo de confiança para a razão referente ao tratamento WW não é apresentado, visto que este tratamento foi considerado como amostra calibradora. Portanto, a comparação da expressão do gene CaACS1-like entre os tratamentos WD e WW não pode ser realizada. Para este mesmo tempo, observa-se que o gene CaACS1-like apresentou expressão significativamente diferente ao se utilizar os tratamentos RW e WD. Para os demais tempos, observa-se que a expressão do gene CaACS1-like não diferiu entre os tratamentos aplicados.

Figura 4.4 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas utilizando os valores das estimativas de  $\Delta\Delta C_q$  obtidas pelo modelo fixo com estrutura de matriz de variâncias e covariâncias ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões



#### 4.4 Modelo misto

Devido à presença de repetições e réplicas no experimento, um novo modelo foi ajustado considerando o efeito de planta (repetição) como aleatório, como sugerido por Steibel et al. (2009). Os efeitos de trat e gene e da interação trat\*gene foram considerados como fixos.

Da mesma maneira como realizado para o modelo fixo que considera todos os valores de  $C_q$ , para o modelo misto ajustado, foram consideradas diferentes matrizes de variâncias e covariâncias residuais para modelar a heterogeneidade de variâncias e a correlação entre as

medidas repetidas. O modelo com matriz de covariância ARH(1) foi o que apresentou melhor ajuste, obtendo menor valor para AIC e BIC.

O resumo dos resultados da análise de variância para o modelo misto sob estrutura ARH(1) da matriz de variâncias e covariâncias residuais é apresentado na Tabela 4.8.

Tabela 4.8 – Resumo da análise de variância para o modelo misto com a estrutura de matriz de variâncias e covariâncias residuais do tipo ARH(1)

Fonte de Variação	Grau de liberdade	Valor-p
Trat	8	<0,0001
Gene	1	<0,0001
Trat*Gene	8	<0,0001

Para a análise utilizando o modelo misto com estrutura ARH(1) da matriz de variâncias e covariâncias residuais, verifica-se pela Tabela 4.8 o efeito significativo da interação entre trat e gene. Como a interação apresentou efeito significativo, foram obtidas as estimativas de  $\Delta\Delta C_q$ , considerando contrastes entre médias, como sugerido por Yuan et al. (2006) e Steibel et al. (2009). As estimativas de  $\Delta\Delta C_q$  são apresentadas na Tabela 4.9. Para cada uma dessas estimativas são apresentados o intervalo de confiança a 95%, o erro padrão e o valor-p do teste t de Student para a hipótese de que  $\Delta\Delta C_q$  é igual a 0.

Tabela 4.9 – Estimativas de  $\Delta\Delta C_q$  obtidas pelo modelo misto com estrutura de matriz de variâncias e covariâncias residuais ARH(1), para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa de  $\Delta\Delta C_q$  são apresentados o erro padrão, o valor-p do teste t de Student e o intervalo de confiança a 95%

Tratamento	Tempo	$\Delta\Delta C_q$	Erro padrão	Valor-p	Intervalo de confiança
RW	2	0,2653	0,27	0,3275	<b>(-0,2685; 0,7991)</b>
WD	2	1,7703	0,267	<0,0001	<i>(1,2424; 2,2981)</i>
WW	2	†	†	†	†
RW	6	0,2737	0,09285	0,0038	<i>(0,0901; 0,4572)</i>
WD	6	0,1753	0,333	0,5995	<b>(-0,4831; 0,8336)</b>
WW	6	0,07351	0,2838	0,796	<b>(-0,4875; 0,6346)</b>
RW	24	1,0899	0,1632	<0,0001	<i>(0,7673; 1,4125)</i>
WD	24	1,15	0,1473	<0,0001	<i>(0,8587; 1,4412)</i>
WW	24	0,7752	0,2193	0,0006	<i>(0,3417; 1,2088)</i>

†Amostra calibradora

Para verificar a existência de diferença significativa entre os genes alvo e referência em cada combinação de tratamento e tempo pode ser utilizado o valor-p do teste t de Student ou o intervalo de confiança obtido. Da mesma forma como especificado no método  $C_q$  comparativo e no modelo fixo, caso opte por realizar a análise utilizando o valor-p, deve-se considerar os genes igualmente expressos para as combinações de tratamento e tempo em que o valor-p é maior

que o nível de significância adotado. Caso deseja-se realizar a análise utilizando o intervalo de confiança, os genes alvo e referência são igualmente expressos nas combinações em que o intervalo para  $\Delta\Delta C_q$  contenha o valor 0. Dessa maneira, verifica-se na Tabela 4.9, que para as combinações tratamento e tempo em que os intervalos de confiança são apresentados em itálico, há evidências de diferença significativa entre os genes alvo e referência. Para as combinações em que os intervalos são apresentados em negrito essa diferença é não significativa.

Utilizando os valores estimados de  $\Delta\Delta C_q$  apresentados na Tabela 4.9, foram obtidas as razões da expressão relativa por meio da fórmula  $2^{-\Delta\Delta C_q}$ . As razões, com os respectivos intervalos de confiança obtidos conforme apresentado na Equação 3.2, são apresentados na Tabela 4.10 e na Figura 4.5.

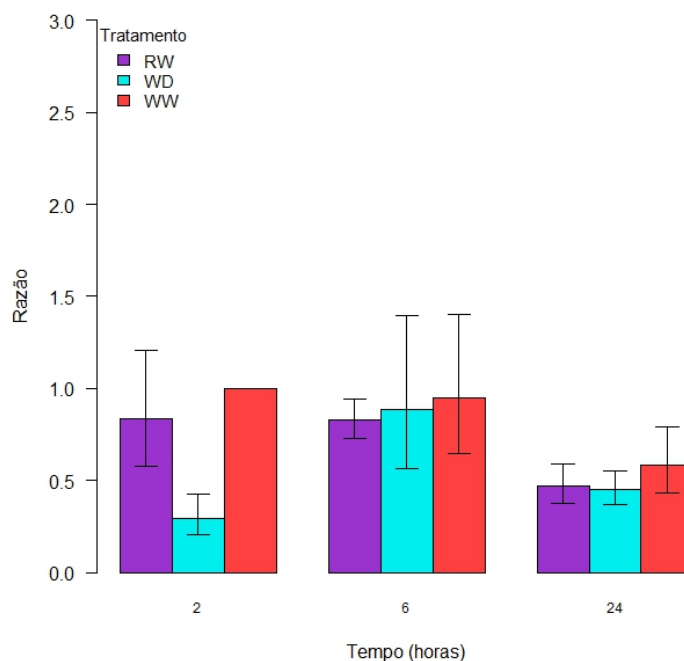
Tabela 4.10 – Estimativas das razões da expressão relativa do gene CaACS1-like utilizando os valores de  $\Delta\Delta C_q$  estimados pelo modelo misto com estrutura de matriz de variâncias e covariâncias residuais ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Para cada estimativa da razão são apresentados o intervalo de confiança a 95%

Tratamento	Tempo	Razão	Intervalo de confiança
RW	2	0,8320	<b>(0,5747; 1,2046)</b>
WD	2	0,2931	<i>(0,2033; 0,4227)</i>
WW	2	†	†
RW	6	0,8272	<i>(0,7284; 0,9394)</i>
WD	6	0,8856	<b>(0,5611; 1,3977)</b>
WW	6	0,9503	<b>(0,6441; 1,4021)</b>
RW	24	0,4698	<i>(0,3756; 0,5875)</i>
WD	24	0,4506	<i>(0,3683; 0,5514)</i>
WW	24	0,5843	<i>(0,4326; 0,7892)</i>

† Amostra calibradora

Assim como nos métodos anteriores, o intervalo de confiança para razão é outro procedimento que pode ser utilizado para verificar se existe diferença na expressão dos genes alvo e referência em cada tratamento. Para as combinações de tratamento e tempo que os intervalos de confiança são apresentados na Tabela 4.10 em negrito, há evidências de que não houve diferença significativa entre os genes CaACS1-like e Ubiquitina, pois para estas combinações o valor 1 encontra-se nos intervalos de confiança. Para os intervalos de confiança apresentados em itálico, há evidências de que para estas combinações de tratamento e tempo a expressão dos genes é diferente, pois para estas combinações o valor 1 não se encontra presente nos intervalos de confiança para a razão. Como já citado, dos 3 procedimentos apresentados, somente um é necessário para verificar a existência de diferença na expressão entre os genes alvo e referência em cada tratamento.

Figura 4.5 – Razões da expressão relativa do gene CaACS1-like analisado na raiz. As barras representam as razões estimadas utilizando os valores das estimativas de  $\Delta\Delta C_q$  obtidas pelo modelo misto com estrutura de matriz de variâncias e covariâncias ARH(1) para os tratamentos reirrigado (RW), déficit de água (WD) e bem irrigado (WW) analisados nos tempos 2, 6 e 24h. Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões



Como especificado, os intervalos de confiança para a razão podem também ser utilizados para comparar a expressão do gene alvo entre os tratamentos aplicados. Caso haja sobreposição dos intervalos de confiança há evidências de que os tratamentos comparados não diferem na expressão do gene. Dessa maneira, pode ser observado na Figura 4.5, que para o tempo 2 a expressão do gene CaACS1-like apresentou diferença significativa ao se comparar os tratamentos RW e WD. Para este mesmo tempo, verifica-se na Figura 4.5 que o intervalo de confiança para a razão ao ser utilizado o tratamento WW não é apresentado, visto que este tratamento foi considerado como a amostra calibradora. Portanto, para o tempo 2 a comparação da expressão do gene CaACS1-like entre os tratamentos WD e WW não pode ser realizada. Para os demais tempos verifica-se que não houve diferença significativa entre os tratamentos.

#### 4.5 Comparação dos métodos aplicados

As razões de expressão relativa para as combinações de tratamento e tempo obtidos pelo modelo de eficiência calibrada, método  $C_q$  comparativo, modelo fixo utilizando a estrutura de variâncias e covariâncias residuais ARH(1) e o modelo misto também sob estrutura de variân-

cias e covariâncias residuais ARH(1) são apresentados na Tabela 4.11 e na Figura 4.6. Para cada razão é apresentado na Tabela 4.11 o intervalo de confiança a 95% e a amplitude deste intervalo. Na Figura 4.6 é apresentado para cada razão o intervalo de confiança a 95%.

Tabela 4.11 – Comparação dos resultados obtidos para as razões e respectivos intervalos de confiança a 95% e amplitudes utilizando o modelo de eficiência calibrada, o método  $C_q$  comparativo, o modelo fixo com estrutura da matriz de variâncias e covariâncias residuais ARH(1) e o modelo misto com estrutura da matriz de variâncias e covariâncias residuais ARH(1)

<b>Método</b>	<b>Tratamento</b>	<b>Tempo</b>	<b>Razão</b>	<b>Intervalo de confiança</b>	<b>Amplitude</b>
<b>Eficiência calibrada</b>	RW	2	0,8206	<b>(-0,1595; 1,8008)</b>	1,9603
	WD	2	0,4020	†	†
	WW	2	1,0031	<b>(0,7653; 1,2409)</b>	0,4758
	RW	6	0,9030	<b>(0,6564; 1,1495)</b>	0,4931
	WD	6	1,1103	<b>(-0,4602; 2,6807)</b>	3,1409
	WW	6	0,9983	<b>(0,1309; 1,8657)</b>	1,7348
	RW	24	0,5173	<b>(-0,0915; 1,1262)</b>	1,2177
	WD	24	0,4887	<i>(0,2039; 0,7736)</i>	0,5697
	WW	24	0,6135	<b>(0,0043; 1,2228)</b>	1,2185
<b>Cq comparativo</b>	RW	2	0,7978	<b>(-0,1651; 1,7607)</b>	1,9258
	WD	2	0,3653	†	†
	WW	2	1,0041	<b>(0,7309; 1,2773)</b>	0,5464
	RW	6	0,8746	<b>(0,6375; 1,1117)</b>	0,4742
	WD	6	1,0813	<b>(-0,5386; 2,7012)</b>	3,2398
	WW	6	0,9916	<b>(-0,0539; 2,0371)</b>	2,091
	RW	24	0,4923	<b>(-0,1002; 1,0848)</b>	1,185
	WD	24	0,4583	<i>(0,1808; 0,7358)</i>	0,555
	WW	24	0,5873	†	†
<b><sup>1</sup>Modelo fixo</b>	RW	2	0,7893	<b>(0,5232; 1,1907)</b>	0,6675
	WD	2	0,2887	<i>(0,2024; 0,4118)</i>	0,2094
	WW	2	††	††	††
	RW	6	0,8834	<b>(0,7656; 1,0194)</b>	0,2538
	WD	6	0,9008	<b>(0,5477; 1,4815)</b>	0,9338
	WW	6	0,9747	<b>(0,6459; 1,4708)</b>	0,8249
	RW	24	0,4632	<i>(0,3561; 0,6024)</i>	0,2643
	WD	24	0,4762	<i>(0,3878; 0,5848)</i>	0,197
	WW	24	0,5917	<i>(0,4176; 0,8384)</i>	0,4208
<b><sup>1</sup>Modelo misto</b>	RW	2	0,832	<b>(0,5747; 1,2046)</b>	0,6299
	WD	2	0,2931	<i>(0,2033; 0,4227)</i>	0,2194
	WW	2	††	††	††
	RW	6	0,8272	<i>(0,7284; 0,9394)</i>	0,211
	WD	6	0,8856	<b>(0,5611; 1,3977)</b>	0,8366
	WW	6	0,9503	<b>(0,6441; 1,4021)</b>	0,758
	RW	24	0,4698	<i>(0,3756; 0,5875)</i>	0,2119
	WD	24	0,4506	<i>(0,3683; 0,5514)</i>	0,1831
	WW	24	0,5843	<i>(0,4326; 0,7892)</i>	0,3566

<sup>1</sup>Modelo ajustado sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1). †Intervalo de confiança e amplitude não obtidos. ††Amostra considerada como calibradora

Os resultados referentes ao modelo fixo utilizando a média das réplicas não são representados na Tabela 4.11 e na Figura 4.6, por este modelo não ter apresentado efeito significativo após a transformação Box-Cox, para a interação entre os fatores trat e gene. Portanto, para todas as combinações de tratamento e tempo presentes na análise, a expressão do gene CaACS1-like não apresentou diferença significativa. A interação trat\*gene para o modelo fixo utilizando o valor  $C_q$  médio foi não significativa possivelmente devido ao fato de não se ter utilizado todas as informações disponíveis do experimento, visto que neste havia a presença de repetições e réplicas. Assim, como observado por Cui e Churchill (2003), Coffey e Cofiel (2005) e Steibel et al. (2009) utilizar a média das réplicas pode prejudicar a análise por não utilizar todas as informações. Ao se ajustar os modelos fixo e misto utilizando os valores de  $C_q$  referentes a todas as réplicas, foi observada a significância da interação entre os fatores gene e tratamento, principal interesse na análise.

Ao se observar a Tabela 4.11 e a Figura 4.6 nota-se que em geral os valores obtidos para a estimativa da razão utilizando o modelo de eficiência calibrada, o método  $C_q$  comparativo e os modelos fixo e misto ajustados sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1) apresentaram resultados similares, o que já era esperado. Ao se comparar somente os resultados obtidos pelo método  $C_q$  comparativo com os resultados obtidos pelo modelo de eficiência calibrada, verifica-se que estes foram semelhantes devido ao fato das eficiências dos genes CaACS1-like(=90%) e Ubiquitina(=99%), definidas para a utilização do modelo de eficiência calibrada, estarem próximas de 100%. Vale ressaltar que o intervalo de aceitação da eficiência para a utilização do método  $C_q$  comparativo, conforme especificado por Lobo e Lobo (2014), deve ser de 90% a 110%. Ao se comparar as razões obtidas pelos métodos que utilizam as eficiências dos genes alvo e referência como sendo iguais a 2, verifica-se que as pequenas diferenças existentes para o valor da razão é devido a existência de parcelas perdidas no experimento. Nos ajustes utilizando os modelos fixo e misto há a correção para o efeito de desbalanceamento, o que não ocorre ao se utilizar o método  $C_q$  comparativo.

Observa-se na Tabela 4.11 que, para as combinações de tratamento e tempo em que os intervalos de confiança são apresentados em **negrito**, há evidências de que os genes CaACS1-like e Ubiquitina se expressaram de maneira similar. Para os intervalos de confiança apresentados em *itálico*, há evidências que a expressão dos genes foi significativamente diferente. Dessa maneira, verifica-se pelo modelo de eficiência calibrada e pelo método  $C_q$  comparativo que os genes são diferentemente expressos somente ao ser utilizado o tratamento WD analisado no



tempo 24. Para os modelos fixos e mistos, ambos sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1), nota-se que entre as 8 combinações de tratamento e tempo analisados, o modelo fixo indicou haver diferença significativa na expressão entre os genes CaACS1-like e Ubiquitina em 4 combinações e o modelo misto em 5 combinações. O modelo misto indicou maior número de diferenças significativas devido a sua maior precisão, pois este ajuste apresentou, na maioria dos casos, menores amplitudes para os intervalos de confiança.

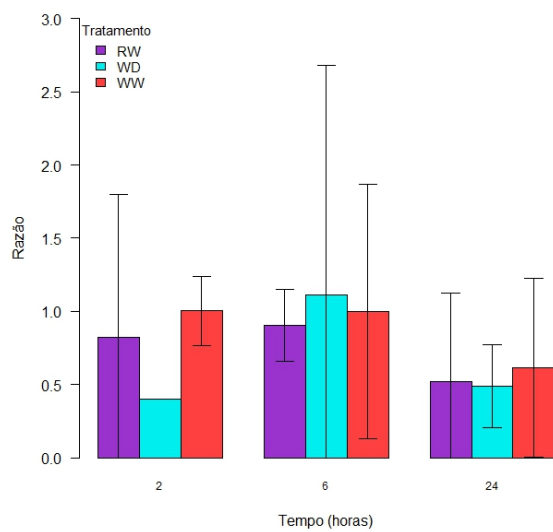
Nota-se na Tabela 4.11 e na Figura 4.6 que ao se utilizar o modelo de eficiência calibrada e o método  $C_q$  comparativo não são apresentados os intervalos de confiança para a razão referente ao tratamento WD analisado no tempo 2, devido ao baixo número de repetições disponíveis, o que impossibilitou a obtenção do valor-p do teste de Shapiro-Wilk, necessário para a verificação da normalidade. Além disso, observa-se que ao se utilizar o método  $C_q$  comparativo, o intervalo de confiança para a razão ao se realizar a análise referente ao tratamento WW no tempo 24 não é apresentado, pois para este tratamento a hipótese de normalidade foi rejeitada. Para os ajustes utilizando os modelos de ANOVA verifica-se que para estas mesmas análises foram apresentados os intervalos de confiança para a razão. Dessa maneira, verifica-se que o modelo de eficiência calibrada e o método  $C_q$  comparativo apresentam como desvantagem a necessidade de se analisar o valor-p para cada combinação de tratamento e tempo e, com isso, a perda de parcelas (repetições) ou a não normalidade dos dados pode fazer com que os intervalos de confiança não possam ser obtidos.

A Figura 4.6 é utilizada para melhor visualizar a comparação da expressão do gene CaACS1-like entre as combinações de tratamentos e tempos aplicados utilizando o modelo de eficiência calibrada (Figura 4.6(a)), método  $C_q$  comparativo (Figura 4.6(b)), modelo fixo sob matriz de variâncias e covariâncias residuais do tipo ARH(1) (Figura 4.6(c)) e modelo misto sob matriz de variâncias e covariâncias residuais do tipo ARH(1) (Figura 4.6(d)). Nesta figura são apresentadas nas barras as razões da expressão relativa e nos segmentos de retas os intervalos de confiança a 95%.

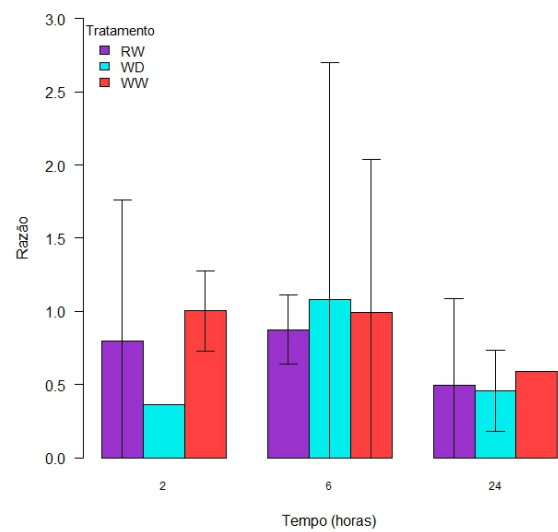
Na Figura 4.6 observa-se que as conclusões obtidas pelo modelo de eficiência calibrada e pelo método  $C_q$  comparativo são similares. Porém, diferentes das obtidas pelos modelos fixo e misto ambos sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1), o que também foi observado por Steibel et al. (2009) ao se comparar o método  $C_q$  comparativo e o modelo misto. Para os métodos tradicionais observa-se que para o tempo 2 a expressão do gene CaACS1-like não diferiu significativamente ao se utilizar os tratamentos RW e WD. Ao

se utilizar os modelos fixo e misto verifica-se que essa diferença é significativa. As conclusões obtidas foram diferentes pois os modelos fixo e misto ajustados obtiveram intervalos de confiança com menores amplitudes comparados aos intervalos obtidos pelo modelo de eficiência calibrada e método  $C_q$  comparativo.

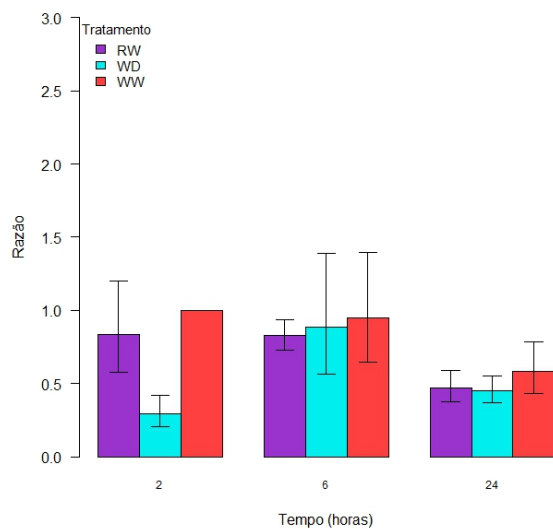
Figura 4.6 – Comparação das razões da expressão relativa do gene CaACS1-like obtidas utilizando os seguintes métodos: modelo de eficiência calibrada; método  $C_q$  comparativo; modelo fixo sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1) e modelo misto sob estrutura da matriz de variâncias e covariâncias residuais do tipo ARH(1). Os segmentos de reta indicam o intervalo de confiança a 95% para estas razões



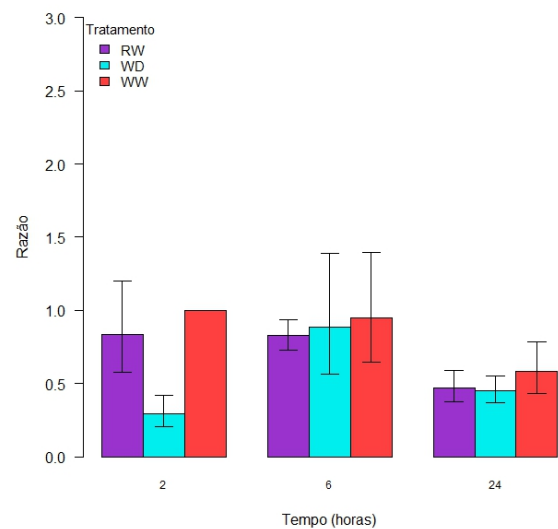
(a) Modelo de eficiência calibrada



(b) Método  $C_q$  Comparativo



(c) Modelo fixo sob estrutura ARH(1) da matriz de variâncias e covariâncias residuais



(d) Modelo misto sob estrutura ARH(1) da matriz de variâncias e covariâncias residuais

Como foi observado na Figura 4.6, os modelos de ANOVA que utilizavam todas as informações presentes no conjunto de dados apresentaram maior precisão nas análises ao serem comparados com as análises realizadas utilizando o modelo de eficiência calibrada e o método  $C_q$  comparativo, por obterem menores amplitudes para os intervalos de confiança. Ao ser comparado o ajuste pelos modelos fixo e misto, observou-se na Tabela 4.11, que na maioria dos casos o modelo misto apresentou menores amplitudes para os intervalos de confiança e, conseqüentemente, indicou maior número de combinações de tratamento e tempo em que ocorre diferença significativa entre os genes CaACS1-like e Ubiquitina. Portanto, o modelo misto apresentou maior precisão nas análises.

O modelo misto apresentou melhor ajuste ao ser comparado com o modelo fixos, devido a adição do termo referente a repetição (planta) que não havia sido incluído no modelo fixo. De acordo com Coffey e Cofiel (2005), quando termos adicionais estão associados ao resultado, a variância residual pode ser substancialmente diminuída e, conseqüentemente, aumenta-se o poder do experimento para observar diferenças nos níveis de expressão.

## 5 CONCLUSÕES

Neste trabalho foram utilizados e comparados quatro métodos envolvendo procedimentos estatísticos para se analisar dados de expressão gênica obtidos pela técnica de PCR em tempo real. Para a base de dados escolhida para as análises foram obtidas as seguintes conclusões:

- Ao se comparar a expressão do gene CaACS1-like entre as combinações de tratamento e tempo aplicados, os resultados obtidos pelo modelo de eficiência e pelo método  $C_q$  comparativo foram similares. Porém, estes resultados não foram equivalentes ao se utilizar os modelos fixo e misto ajustados sob estrutura da matriz de variâncias e covariâncias residuais do tipo autorregressiva heterogênea de primeira ordem (ARH(1)). Para os ajustes pelos modelos de ANOVA, as conclusões em relação a comparação da expressão do gene alvo entre os tratamentos foram as mesmas e mais precisas que as conclusões obtidas pelo modelo de eficiência calibrada e pelo método  $C_q$  comparativo .
- O modelo misto foi considerado o que apresentou melhores resultados por obter, em geral, menores amplitudes para os intervalos de confiança para a razão, apresentando maior precisão nas análises em comparação com os demais métodos aplicados. Consequentemente, por apresentar menores intervalos de confiança, o modelo misto conseguiu detectar maior número de diferença significativa na expressão dos genes CaACS1-like e Ubiquitina para as combinações de tratamento e tempo presentes na análise.

## REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, New York, v. 19, n. 6, p. 716–723, 1974.
- ALMEIDA, A. de; ELIAN, S.; NOBRE, J. Modificações e alternativas aos testes de levene e de brown e forsythe para igualdade de variâncias e médias. **Revista Colombiana de Estadística**, Bogotá, v. 31, n. 2, p. 241–260, 2008.
- ANDRADE, G. M. de. **Estatística Geral e Aplicada**. 4. ed. São Paulo: Editora Atlas SA, 2011.
- APPLIED BIOSYSTEMS. **Guide to performing relative quantitation of gene expression using Real-Time quantitative PCR**. Foster City: [s.n.], 2004. Disponível em: <[http://www.gu.se/digitalAssets/1125/1125331\\_ABI\\_-\\_Guide\\_Relative\\_Quantification\\_using\\_realtme\\_PCR.pdf](http://www.gu.se/digitalAssets/1125/1125331_ABI_-_Guide_Relative_Quantification_using_realtme_PCR.pdf)>. Acesso em: 23 ago. 2016.
- BARBIN, D. **Planejamento e análise estatística de experimentos agrônômicos**. Araporangas: Midas, 2003.
- BAUTISTA, E.; QUARESMA, E.; SANTANA, M. Modelos lineares mistos para análise de dados longitudinais provenientes de ensaios agrônômicos. **Revista da Estatística da Universidade Federal de Ouro Preto**, Ouro Preto, v. 3, n. 3, p. 649–653, 2014.
- BENÍTEZ, R. et al. Effects of dietary fat saturation on fatty acid composition and gene transcription in different tissues of iberian pigs. **Meat science**, Oxford, v. 102, p. 59–68, 2015.
- BERTOLDO, J. et al. Teste de comparação de médias: dificuldades e acertos em artigos científicos. **Current Agricultural Science and Technology**, Tehran, v. 13, n. 4, p. 441–447, 2007.
- BOX, G. E.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, p. 211–252, 1964.
- BUSATO, K. et al. Expression of genes related to the regulation of muscle protein turnover in Angus and Nellore bulls. **Journal of animal science**, Champaign, v. 94, n. 4, p. 1472–1481, 2016.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. Pacific Grove: Duxbury Pacific Grove, 2002. v. 2.
- CHEN, K.; FESSEHAIE, A.; ARORA, R. Selection of reference genes for normalizing gene expression during seed priming and germination using qPCR in *Zea mays* and *Spinacia oleracea*. **Plant Molecular Biology Reporter**, New York, v. 30, n. 2, p. 478–487, 2012.
- CHUNG, S.-H.; PEARN, W.; YANG, Y.-S. A comparison of two methods for transforming non-normal manufacturing data. **The International Journal of Advanced Manufacturing Technology**, Bedford, v. 31, n. 9-10, p. 957–968, 2007.
- COFFEY, C. S.; COFIEL, S. S. Parametric linear models. In: ALLISON, D. B. et al. (Ed.). **DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments (Biostatistics)**. 1. ed. New York: Chapman and Hall/CRC, 2005.

- CUI, X.; CHURCHILL, G. A. Statistical tests for differential expression in cDNA microarray experiments. **Genome biology**, London, v. 4, n. 4, p. 210, 2003.
- FAIBISH, D.; SUZUKI, M.; BARTLETT, J. Appropriate real-time PCR reference genes for fluoride treatment studies performed in vitro or in vivo. **Archives of oral biology**, Oxford, v. 62, p. 33–42, 2016.
- FERREIRA, D. F. **Estatística Básica**. 2. ed. Lavras: UFLA, 2009.
- GONI, R.; GARCÍA, P.; FOISSAC, S. The qPCR data statistical analysis. **Integromics White Paper**, London, p. 1–9, 2009.
- GONZÁLEZ-CALVO, L. et al. Effect of finishing period length with  $\alpha$ -tocopherol supplementation on the expression of vitamin E-related genes in the muscle and subcutaneous fat of light lambs. **Gene**, Amsterdam, v. 552, n. 2, p. 225–233, 2014.
- HEIBERGER, R. M.; HOLLAND, B. **Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS**. Berlim: [s.n.], 2004.
- HUANG, H.; XU, Y.; CHENG, C. Detection of alternative splicing during epithelial-mesenchymal transition. **Journal of visualized experiments: JoVE**, Boston, n. 92, 2014.
- KUBISTA, M. et al. The real-time polymerase chain reaction. **Molecular aspects of medicine**, Oxford, v. 27, n. 2, p. 95–125, 2006.
- LAU, J. A.; LENNON, J. T. Evolutionary ecology of plant–microbe interactions: soil microbial structure alters selection on plant traits. **New Phytologist**, London, v. 192, n. 1, p. 215–224, 2011.
- LIMA, A. A. **Ethylene regulation under different watering conditions and its possible involvement in coffee (*Coffea arabica* L.) flowering**. Tese (Doutorado) — Universidade Federal de Lavras, Lavras, 2015.
- LITTELL, R. C. et al. **SAS for mixed models**. 2. ed. North Carolina: SAS institute, 2006.
- LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta C_T}$  Method. **Methods**, San Diego, v. 25, n. 4, p. 402–408, 2001.
- LOBO, A. M. B. O.; LOBO, R. N. B. Considerações Estatísticas na Análise de Dados de Expressão gênica Gerados pela Técnica de RT-qPCR. **Embrapa Caprinos e Ovinos-Documentos (INFOTECA-E)**, Sobral, 2014.
- LÚCIO, A. D. et al. Transformação box-cox em experimentos com pimentão em ambiente protegido. **Horticultura brasileira**, Brasília, v. 29, n. 1, 2011.
- MCCULLOCH, R. et al. Progression of gene expression changes following a mechanical injury to articular cartilage as a model of early stage osteoarthritis. **Arthritis**, Malden, v. 2014, 2014.
- NASCIMENTO, S.; SUAREZ, E. R.; PINHAL, M. A. d. S. Tecnologia de PCR e RT-PCR em tempo real e suas aplicações na área médica. **RBM rev. bras. med**, Rio de Janeiro, v. 67, n. supl. 10, 2010.

- NOGUEIRA, M. C. S.; CORRENTE, J. E. Decomposição da interação tripla significativa utilizando o comando contrast do PROC GLM do SAS, aplicado ao modelo de classificação tripla para dados balanceados. **Bragantia**, Campinas, v. 59, n. 1, p. 109–115, 2000.
- NOLAN, T.; HANDS, R. E.; BUSTIN, S. A. Quantification of mRNA using real-time RT-PCR. **Nature protocols**, London, v. 1, n. 3, p. 1559–1582, 2006.
- PABINGER, S. et al. A survey of tools for the analysis of quantitative PCR (qPCR) data. **Biomolecular Detection and Quantification**, Munich, v. 1, n. 1, p. 23–33, 2014.
- PAOLACCI, A. R. et al. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. **BMC molecular biology**, London, v. 10, n. 1, 2009.
- PERECIN, D.; CARGNELUTTI FILHO, A. Efeitos por comparações e por experimento em interações de experimentos fatoriais. **Ciência e Agrotecnologia**, Lavras, v. 32, n. 1, p. 68–72, 2008.
- PFAFFL, M. W. A new mathematical model for relative quantification in real-time RT-PCR. **Nucleic acids research**, London, v. 29, n. 9, p. 2003–2007, 2001.
- POSTOLLEC, F. et al. Recent advances in quantitative PCR (qPCR) applications in food microbiology. **Food microbiology**, London, v. 28, n. 5, p. 848–861, 2011.
- PRADEEPA, M. M. et al. Histone h3 globular domain acetylation identifies a new class of enhancers. **Nature genetics**, New York, 2016.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>.
- RIEU, I.; POWERS, S. J. Real-time quantitative RT-PCR: design, calculations, and statistics. **The Plant Cell**, Rockville, v. 21, n. 4, p. 1031–1033, 2009.
- SCHMITTGEN, T. D.; LIVAK, K. J. Analyzing real-time pcr data by the comparative ct method. **Nature protocols**, London, v. 3, n. 6, p. 1101–1108, 2008.
- SCHWARZ, G. Estimating the dimension of a model. **The annals of statistics**, Ithaca, v. 6, n. 2, p. 461–464, 1978.
- SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. New Jersey: John Wiley & Sons, 2009. v. 391.
- SMITHSON, M. **Confidence Intervals: Quantitative Applications in the Social Sciences**. 1. ed. CA: Sage Publications, Inc, 2003.
- STEIBEL, J. P. et al. A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. **Genomics**, San Diego, v. 94, n. 2, p. 146–152, 2009.
- STORCK, L. et al. Análise de covariância para melhoria da capacidade de discriminação em ensaios de cultivares de milho. **Pesquisa Agropecuária Brasileira**, Rio de Janeiro, v. 35, n. 7, p. 1311–1316, 2000.
- TICHOPAD, A. et al. Experiment Design, Data Management, and Univariate Statistical Analysis of Gene-Expression Data Obtained by Real-Time Quantitative PCR. In: **PCR Technology: Current Innovations, Third Edition**. Boca Raton: CRC Press, 2013. p. 245–258.

- TORAL, F. L. B.; ALENCAR, M. M. de; FREITAS, A. R. de. Arranjos para efeitos fixos e estruturas de (co) variâncias residuais para análises de medidas repetidas do peso de bovinos da raça canchim. **R. Bras. Zootec**, Cuiabá, v. 35, n. 5, p. 1951–1958, 2006.
- TORMAN, V. B. L.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. **Revista HCPA**, Porto Alegre, v. 32, n. 2, p. 227–234, 2012.
- UDVARDI, M. K.; CZECHOWSKI, T.; SCHEIBLE, W.-R. Eleven golden rules of quantitative RT-PCR. **The Plant Cell**, London, v. 20, n. 7, p. 1736–1737, 2008.
- VANDESOMPELE, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome biology**, London, v. 3, n. 7, p. 1–12, 2002.
- VANGUILDER, H. D.; VRANA, K. E.; FREEMAN, W. M. Twenty-five years of quantitative PCR for gene expression analysis. **Biotechniques**, London, v. 44, n. 5, p. 619–626, 2008.
- WEST, B. T.; WELCH, K. B.; GALECKI, A. T. **Linear mixed models: a practical guide using statistical software**. 2. ed. New York: CRC Press, 2014.
- WONG, M. L.; MEDRANO, J. F. Real-time PCR for mRNA quantitation. **Biotechniques**, London, v. 39, n. 1, p. 75–85, 2005.
- YUAN, J. S. et al. Statistical analysis of real-time PCR data. **BMC bioinformatics**, London, v. 7, n. 85, 2006.



## APÊNDICE A – Rotina no programa R

```
#####
##### Modelo de eficiência calibrada #####
#####
#Eficiências dos genes:
#EficAlvo=1.90
#EficUbq=1.99
#Lendo os dados
rm(list=ls())
setwd("C:\\Users\\Andrezza\\Desktop\\CaACS1like")
dados1<- read.table("CaACS1likeRaiz.txt", head=T, dec=".")
head(dados1)
#Abrindo os pacotes necessários
library(lubridate)
library(dplyr)
#Verificando a estrutura dos dados
str(dados1)
#Calculando a média das réplicas
#O Cq agora passará a ser a média das réplicas
Cq=with(dados1,tapply(Cq,interaction(Tratamento,Tempo,Gene,Repeticao),
                    mean, na.rm=TRUE))
Tabela=expand.grid(unique(dados1$Tratamento), unique(dados1$Tempo),
                    unique(dados1$Gene), unique(dados1$Repeticao))
names(Tabela)[1:4]=c("Tratamento", "Tempo", "Gene", "Repeticao" )
#Verificando como esta ordenado os elementos de Cq
Cq=as.data.frame(Cq)
head(Cq)
#Ordenando a Tabela para ficar igual a ordem que saiu o valor de Cq
Tabela=Tabela[order(Tabela$Repeticao, Tabela$Gene, Tabela$Tempo,
                    Tabela$Tratamento),]
#Adicionando o valor de Cq na Tabela anterior
Tabela2= cbind(Tabela,Cq)
```

```

#Amostra Calibradora
#Média das Repetições na amostra CALIBRADORA
#Média Repetiçao Cq no gene Referência na amostra Calibradora
CqRepUbqCont=Tabela2$Cq[Tabela2$Gene=="Ubq" & Tabela2$Tratamento=="WW"
      & Tabela2$Tempo=="2"]
MediaRepeticaoCqUbqCont=mean(c(CqRepUbqCont), na.rm=T)
#Média Repetiçao Cq no gene Alvo na amostra Calibradora
CqRepAlvoCont=Tabela2$Cq[Tabela2$Gene=="CaACS1like" &
      Tabela2$Tratamento=="WW" & Tabela2$Tempo=="2"]
MediaRepeticaoCqAlvoCont=mean(c(CqRepAlvoCont), na.rm=T)
#Separando a tabela em Alvo e Referência
TabelaAlvo=filter(Tabela2, Tabela2$Gene=="CaACS1like" )
TabelaRef=filter(Tabela2, Tabela2$Gene=="Ubq" )
#Separando os Cq's Alvo e referência na amostra tratada
CqAlvoTrat=TabelaAlvo$Cq
CqRefTrat=TabelaRef$Cq
#Incluindo as eficiências dos genes:
EficAlvo=1.90
EficUbq=1.99
#Calculando Delta Cq do Alvo e referência
DCqAlvo=MediaRepeticaoCqAlvoCont-CqAlvoTrat
DCqRef=MediaRepeticaoCqUbqCont-CqRefTrat
#Obtendo a Razão
Razao=(EficAlvo^DCqAlvo)/(EficUbq^DCqRef)
Tabela0=expand.grid(unique(Tabela2$Tratamento),
unique(Tabela2$Tempo), unique(Tabela2$Repeticao))
names(Tabela0)[1:3]=c("Tratamento", "Tempo" , "Repeticao")
Tabela1=cbind(Tabela0, Razao)
Razao=with(Tabela1,tapply(Razao,interaction(Tratamento,Tempo),
      mean, na.rm=TRUE))
#Teste de Shapiro-Wilk (Normalidade) para Razão
#(Script apresentado somente para o tratamento RW no tempo 2)

```

```

RazaoRaizRWT2=Tabela1$Razao[Tabela1$Tratamento=="RW" & Tabela1$Tempo=="2"]
normRazaoRaizRWT2=shapiro.test(RazaoRaizRWT2)
ValorPNormRazaoRRWT2=normRazaoRaizRWT2$p.value
#Calculando o erro padrão da razão
SDRazao=with(Tabela1,tapply(Razao,interaction(Tratamento,Tempo),
                             sd, na.rm=TRUE))
EPRazao=SDRazao/sqrt(3)
EPRazao=round(EPRazao, 4)
#Obtendo IC para a Razao
#IC Razão 95% Supondo normalidade
ttab=qt(c(.975), df=2)
ICsupRazao=Razao+(ttab*EPRazao)
ICsupRazao=round(ICsupRazao, 4)
ICinfRazao=Razao-(ttab*EPRazao)
ICinfRazao=round(ICinfRazao, 4)
#Retirando o intervalo de confiança para a Razão para o tratamento WD
# no tempo 2 (Valor-p do teste de Shapiro-Wilk não obtido)
ICinfRazao[2]=NA
ICsupRazao[2]=NA
#Tabela Final
Tabelafinal=expand.grid(unique(Tabela1$Tratamento),unique(Tabela1$Tempo))
names(Tabelafinal)[1:2]=c("Tratamento", "Tempo")
Tabelafinal=cbind(Tabelafinal, Razao, EPRazao, ICinfRazao,ICsupRazao)
#Construindo o gráfico de barras com IC
tabelamedias <- tapply(Tabelafinal$Razao, list(Tabelafinal$Tratamento,
                                             Tabelafinal$Tempo),function(x) c(x = x))
tabelaICinf <- tapply(Tabelafinal$ICinfRazao, list(Tabelafinal$Tratamento,
                                                  Tabelafinal$Tempo),function(x) c(x = x))
tabelaICsup <- tapply(Tabelafinal$ICsupRazao, list(Tabelafinal$Tratamento,
                                                  Tabelafinal$Tempo), function(x) c(x = x))
graficobarrasCaACS1likeRaizIC <-barplot(height = tabelamedias,
                                          beside = TRUE, las = 1, col=c("darkorchid3", "cyan2", "brown1"),

```

```

ylim = c(0, 3), cex.names = 0.75, ylab = "Razão",
xlab = "Tempo (horas)", border = "black", axes = TRUE,
legend.text = TRUE, args.legend = list(title = "Tratamento",
x = "topleft", bty = "n", cex = .9)
arrows(graficobarrasCaACS1likeRaizIC, tabelaICinf,
      graficobarrasCaACS1likeRaizIC, tabelaICsup, length=0.1, angle=90,
      code=3)
#####
##### Método Cq Comparativo #####
#####
#Lendo os dados
rm(list=ls())
setwd("C:\\Users\\Andrezza\\Desktop\\CaACS1like")
dados1<- read.table("CaACS1likeRaiz.txt", head=T, dec=".")
#Abrindo os pacotes necessários
library(lubridate)
library(dplyr)
#Verificando a estrutura dos dados
str(dados1)
#Calculando a média das réplicas
#O Cq agora passará a ser a média das réplicas
Cq=with(dados1,tapply(Cq,interaction(Tratamento,Tempo,Gene, Repeticao),
      mean, na.rm=TRUE))
Tabela=expand.grid(unique(dados1$Tratamento), unique(dados1$Tempo),
      unique(dados1$Gene), unique(dados1$Repeticao))
names(Tabela)[1:4]=c("Tratamento", "Tempo", "Gene", "Repeticao" )
#Verificando como esta ordenado os elementos de Cq
Cq=as.data.frame(Cq)
head(Cq)
#Ordenando a Tabela para ficar igual a ordem que saiu o valor de Cq
Tabela=Tabela[order(Tabela$Repeticao, Tabela$Gene, Tabela$Tempo,
      Tabela$Tratamento),]

```

```

head(Tabela)
#Adicionando o valor de Cq na Tabela anterior
Tabela2= cbind(Tabela,Cq)
#Amostra Calibradora
#Média das Repetições na amostra CALIBRADORA
#Média Repetiçao Cq no gene Referência na amostra Calibradora
CqRepUbqCont=Tabela2$Cq[Tabela2$Gene=="Ubq" & Tabela2$Tratamento=="WW" &
    Tabela2$Tempo=="2"]
MediaRepeticaoCqUbqCont=mean(c(CqRepUbqCont), na.rm=T)
#Média Repetiçao Cq no gene Alvo na amostra Calibradora
CqRepAlvoCont=Tabela2$Cq[Tabela2$Gene=="CaACS1like" &
    Tabela2$Tratamento=="WW" & Tabela2$Tempo=="2"]
MediaRepeticaoCqAlvoCont=mean(c(CqRepAlvoCont), na.rm=T)
#Separando a tabela em Alvo e Referência
TabelaAlvo=filter(Tabela2, Tabela2$Gene=="CaACS1like" )
TabelaRef=filter(Tabela2, Tabela2$Gene=="Ubq" )
#Separando os Cq's Alvo e referência na amostra tratada
CqAlvoTrat=TabelaAlvo$Cq
CqRefTrat=TabelaRef$Cq
#Calculando os Delta Cq's Alvo e Referência
DeltaCtAlvo=MediaRepeticaoCqAlvoCont-CqAlvoTrat
DeltaCtRef=MediaRepeticaoCqUbqCont-CqRefTrat
# Calculando Delta Delta Cq e a Razão
DDCq0=DeltaCtRef-DeltaCtAlvo
Razao0=2^-(DDCq0)
Tabela0=expand.grid(unique(Tabela2$Tratamento), unique(Tabela2$Tempo),
    unique(Tabela2$Repeticao))
names(Tabela0)[1:3]=c("Tratamento", "Tempo", "Repeticao")
Tabela1=cbind(Tabela0, DDCq0, Razao0)
DDCq=with(Tabela1,tapply(DDCq0, interaction(Tratamento,Tempo), mean,
    na.rm=TRUE))
DDCq=round(DDCq, 4)

```

```

Razao=with(Tabela1,tapply(Razao0,interaction(Tratamento,Tempo), mean,
                           na.rm=TRUE))
Razao=round(Razao, 4)
#Teste de Shapiro-Wilk (Normalidade) para DDCq
      #(Script apresentado somente para o tratamento RW no tempo 2)
DDCqRaizRWT2=Tabela1$DDCq0[Tabela1$Tratamento=="RW" & Tabela1$Tempo=="2"]
normDDCqRaizRWT2=shapiro.test(DDCqRaizRWT2)
ValorPNormDDCqRRWT2=normDDCqRaizRWT2$p.value
#Teste de Shapiro-Wilk (Normalidade) para Razão
      #(Script apresentado somente para o tratamento RW no tempo 2)
RazaoRaizRWT2=Tabela1$Razao0[Tabela1$Tratamento=="RW" &
                              Tabela1$Tempo=="2"]
normRazaoRaizRWT2=shapiro.test(RazaoRaizRWT2)
ValorPNormRazaoRRWT2=normRazaoRaizRWT2$p.value
#Calculando o EP de DDCq e da Razao
SDDDCq=with(Tabela1,tapply(DDCq0,interaction(Tratamento,Tempo),
                           sd, na.rm=TRUE))
EPDDCq=SDDDCq/sqrt(3)
EPDDCq=round(EPDDCq, 4)
SDRazao=with(Tabela1,tapply(Razao0,interaction(Tratamento,Tempo),
                           sd, na.rm=TRUE))
EPRazao=SDRazao/sqrt(3)
EPRazao=round(EPRazao, 4)
#Obtendo IC para DDCq
ttab=qt(c(.975), df=2)
ICsupDDCq=DDCq+(ttab*EPDDCq)
ICsupDDCq=round(ICsupDDCq, 4)
ICinfDDCq=DDCq-(ttab*EPDDCq)
ICinfDDCq=round(ICinfDDCq, 4)
#Obtendo IC para a Razao
      #IC Razão 95% Supondo normalidade
ttab=qt(c(.975), df=2)

```

```

ICsupRazao=Razao+(ttab*EPRazao)
ICsupRazao=round(ICsupRazao, 4)
ICinfRazao=Razao-(ttab*EPRazao)
ICinfRazao=round(ICinfRazao, 4)
#Retirando o intervalo de confiança para a Razão para o tratamento WD no
      # tempo 2 e WW no tempo 24 (não normalidade ou valor-p não obtido)
ICinfRazao[2]=NA
ICsupRazao[2]=NA
ICinfRazao[9]=NA
ICsupRazao[9]=NA
#Aplicando o Teste t de Student
# CALIBRADORA
#Obtendo DeltaCq para a amostra calibradora
DeltaCqCont=CqRepAlvoCont-CqRepUbqCont
#Verificando as pressuposições do teste t para a amostra calibradora
#Teste de normalidade
norm.test.DCqCont=shapiro.test(DeltaCqCont)
norm.test.DCqCont
#Amostra tratada
      #(Script apresentado somente para o tratamento RW no tempo 2)
CqAlvoRWT2=Tabela2$Cq[Tabela2$Tratamento=="RW" & Tabela2$Tempo=="2" &
      Tabela2$Gene=="CaACS1like"]
CqRefRWT2=Tabela2$Cq[Tabela2$Tratamento=="RW" & Tabela2$Tempo=="2" &
      Tabela2$Gene=="Ubq"]
#Obtendo DeltaCq para a amostra tratada
DeltaCqRWT2=CqAlvoRWT2-CqRefRWT2
#Verificando as pressuposições do teste t
#Teste de normalidade
norm.test.DCqRWT2=shapiro.test(DeltaCqRWT2)
norm.test.DCqRWT2
#Igualdade de Variâncias:
var.test = var.test(DeltaCqCont,DeltaCqRWT2,conf.level=0.95)

```

```

valor.p.varRWT2 = var.test$p.value
valor.p.varRWT2
#Aplicação do teste t
t.testeRWT2=t.test(DeltaCqWWT2,DeltaCqRWT2, var.equal=T, alt="two.sided",
paired = F)
valor.p.teste.t.RWT2=t.testeRWT2$p.value
#Tabela Final
Tabelafinal=expand.grid(unique(Tabelafinal$Tratamento),
unique(Tabelafinal$Tempo))
names(Tabelafinal)[1:2]=c( "Tratamento", "Tempo" )
Tabelafinal=cbind(Tabelafinal, DDCq, EPDDCq, ICinfDDCq,ICsupDDCq,
Razao, EPRazao, ICinfRazao,ICsupRazao)
#Construindo o gráfico de barras com o IC da razão
tabelamedias <- tapply(Tabelafinal$Razao, list(Tabelafinal$Tratamento,
Tabelafinal$Tempo), function(x) c(x = x))
tabelaICinf <- tapply(Tabelafinal$ICinfRazao, list(Tabelafinal$Tratamento,
Tabelafinal$Tempo), function(x) c(x = x))
tabelaICsup <- tapply(Tabelafinal$ICsupRazao, list(Tabelafinal$Tratamento,
Tabelafinal$Tempo), function(x) c(x = x))
graficobarrasCaACS1likeRaizIC <- barplot(height = tabelamedias,
beside = TRUE, las = 1, col=c("darkorchid3", "cyan2", "brown1"),
ylim = c(0, 3), cex.names = 0.75, ylab = "Razão",
xlab = "Tempo (horas)", border = "black", axes = TRUE,
legend.text = TRUE, args.legend = list(title = "Tratamento",
x = "topleft", bty = "n", cex = .9))
arrows(graficobarrasCaACS1likeRaizIC, tabelaICinf,
graficobarrasCaACS1likeRaizIC, tabelaICsup, length=0.1, angle=90,
code=3)
#####
##### ANOVA FIXO #####
#####
#Lendo os dados

```



```
rm(list=ls())
setwd("C:\\Users\\Andrezza\\Desktop\\CaACS1like")
dados1<- read.table("CaACS1likeRaiz.txt", head=T, dec=".")
#Abrindo os pacotes necessários
library(lubridate)
library(dplyr)
library(MASS)
#Verificando a estrutura dos dados
str(dados1)
#Transformando o tempo em fator
dados1$Tempo=as.factor(dados1$Tempo)

##### Ajustando o modelo fixo com a média das réplicas #####

#Calculando a média das réplicas
#O Cq agora passará a ser a média das réplicas
Cq=with(dados1,tapply(Cq,interaction(Tratamento,Tempo,Gene, Repeticao),
                    mean, na.rm=TRUE))
Tabela=expand.grid(unique(dados1$Tratamento), unique(dados1$Tempo),
                    unique(dados1$Gene), unique(dados1$Repeticao))
names(Tabela)[1:4]=c("Tratamento", "Tempo", "Gene", "Repeticao" )
#Verificando como esta ordenado os elementos de Cq
Cq=as.data.frame(Cq)
head(Cq)
#Ordenando a Tabela para ficar igual a ordem que saiu o valor de Cq
Tabela=Tabela[order(Tabela$Repeticao, Tabela$Gene, Tabela$Tempo,
                    Tabela$Tratamento),]
#Adicionando o valor de Cq na Tabela anterior
Tabela2= cbind(Tabela,Cq)
#Criando o fator Trat (combinação de tratamento e tempo)
Tabela2$Trat=paste(Tabela2$Tratamento, Tabela2$Tempo, sep="")
#Ajuste do modelo
```

```

#Anova tipo III (dados desbalanceados)
options(contrasts = c("contr.sum","contr.poly"))
modelogene1 <- lm(Cq~Trat*Gene, data=Tabela2)
drop1(modelogene1, .~., test="F")
#Verificando as pressuposições do modelo
#Teste de Shapiro-Wilk
shapiro.test(residuals(modelogene1))
#Homocedasticidade:
Tabela2$TG=paste(Tabela2$Trat, Tabela2$Gene, sep="")
bartlett.test(Tabela2$Cq, Tabela2$TG)
#Transformação de Box-Cox
boxcox(Cq~Trat*Gene, data=Tabela2, plotit=T, lam=seq(-8,4,0.05))
#Valor zero esta incluído no IC o que sugere realizar transformação log
Tabela2$Cq.t.log=log(Tabela2$Cq)
#Ajuste do modelo com média após transformação
options(contrasts = c("contr.sum","contr.poly"))
modelogene2 <- lm(Cq.t.log~Trat*Gene, data=Tabela2)
drop1(modelogene2, .~., test="F")
#Verificando as pressuposições do modelo após transformação
#Teste de Shapiro-Wilk (Normalidade)
shapiro.test(residuals(modelogene2))
#Homocedasticidade:
Tabela2$TG=paste(Tabela2$Trat, Tabela2$Gene, sep="")
bartlett.test(Tabela2$Cq.t.log, Tabela2$TG)

##### Ajustando o modelo fixo com todos os valores de Cq #####
##### presentes no conjunto de dados #####
rm(list=ls())
setwd("C:\\Users\\Andrezza\\Desktop\\CaACS1like")
dados1<- read.table("CaACS1likeRaiz.txt", head=T, dec=".")
#Abrindo os pacotes necessários
library(lubridate)

```

```
library(dplyr)
library(MASS)
#Verificando a estrutura dos dados
str(dados1)
#Transformando o tempo em fator
dados1$Tempo=as.factor(dados1$Tempo)
#Criando o fator Tratamento*Tempo
dados1$Trat=paste(dados1$Tratamento, dados1$Tempo, sep="")
dados1$Trat=as.factor(dados1$Trat)
#Ajuste do modelo:
options(contrasts = c("contr.sum", "contr.poly"))
modelogene3 <- lm(Cq~Trat*Gene, data=dados1)
drop1(modelogene3, .~., test="F")
#Verificando as pressuposições do modelo após transformação
#Teste de Shapiro-Wilk (Normalidade)
shapiro.test(residuals(modelogene3))
#Homocedasticidade:
dados1$TG=paste(dados1$Trat, dados1$Gene, sep="")
bartlett.test(dados1$Cq, dados1$TG)

##Obs: Ajuste utilizando a matriz de variâncias e covariâncias residual
      ARH(1) para o modelo fixo e misto e obtenção dos valores
      referentes as estimativas de DDCq e respectivos EP e IC
      foram realizados utilizando o software SAS ###
      ##(Rotina apresentada no apêndice B) ##

###Cálculo das razões e respectivos IC para o modelo de ANOVA FIXO

#Obtendo as estimativas das razões
      #(utilizando as estimativas de DDCq obtidas no SAS)
Razao=2^- (dados1$DDCq)
Razao=round(Razao, 4)
```

```

#Obtendo o intervalo de confiança para a razão
      #(utilizando os valores de EP de DDCq e GL obtidos no SAS)
ttab=qt(c(.975), df=15)
ICsupRazao=2^-(dados1$DDCq-(ttab*dados1$EPDDCq))
ICsupRazao=round(ICsupRazao, 4)
ICinfRazao=2^-(dados1$DDCq+(ttab*dados1$EPDDCq))
ICinfRazao=round(ICinfRazao, 4)
#Tabela Final Razao
Tabelafinal=expand.grid(unique(dados1$Tratamento), unique(dados1$Tempo))
names(Tabelafinal)[1:2]=c("Tratamento", "Tempo")
TabelaFinal=cbind(Tabelafinal, Razao, ICinfRazao, ICsupRazao)
#Construindo o gráfico de barras com o IC para a Razao
tabelamedias <- tapply(TabelaFinal$Razao, list(TabelaFinal$Tratamento,
      TabelaFinal$Tempo), function(x) c(x = x))
tabelaICinf<- tapply(TabelaFinal$ICinfRazao, list(TabelaFinal$Tratamento,
      TabelaFinal$Tempo), function(x) c(x = x))
tabelaICsup<- tapply(TabelaFinal$ICsupRazao, list(TabelaFinal$Tratamento,
      TabelaFinal$Tempo), function(x) c(x = x))
graficobarrasCaACS1likeRaizIC <- barplot(height = tabelamedias,
      beside = TRUE, las = 1, col=c("darkorchid3", "cyan2", "brown1"),
      ylim = c(0, 3), cex.names = 0.75, ylab = "Razão",
      xlab = "Tempo (horas)", border = "black", axes = TRUE,
      legend.text = TRUE, args.legend = list(title = "Tratamento",
      x = "topleft", bty = "n", cex = .9))
arrows(graficobarrasCaACS1likeRaizIC, tabelaICinf,
      graficobarrasCaACS1likeRaizIC, tabelaICsup, length=0.1, angle=90,
      code=3)

###Cálculo das razões e respectivos IC para o modelo misto

#Obtendo as estimativas das razões
      #(utilizando as estimativas de DDCq obtidas no SAS)

```

```

Razao=2^-(dados1$DDCq)
Razao=round(Razao, 4)
#Obtendo o intervalo de confiança para a razão
      #(utilizando os valores de EP de DDCq e GL obtidos no SAS)
ttab=qt(c(.975), df=139)
ICsupRazao=2^-(dados1$DDCq-(ttab*dados1$EPDDCq))
ICsupRazao=round(ICsupRazao, 4)
ICinfRazao=2^-(dados1$DDCq+(ttab*dados1$EPDDCq))
ICinfRazao=round(ICinfRazao, 4)
#Tabela Final Razao
Tabelafinal=expand.grid(unique(dados1$Tratamento),
unique(dados1$Tempo))
names(Tabelafinal)[1:2]=c( "Tratamento", "Tempo" )
TabelaFinal=cbind(Tabelafinal, Razao, ICinfRazao,ICsupRazao)
#Construindo o gráfico de barras com o IC para a Razao
tabelamedias <- tapply(TabelaFinal$Razao, list(TabelaFinal$Tratamento,
      TabelaFinal$Tempo), function(x) c(x = x))
tabelaICinf<- tapply(TabelaFinal$ICinfRazao, list(TabelaFinal$Tratamento,
      TabelaFinal$Tempo), function(x) c(x = x))
tabelaICsup<- tapply(TabelaFinal$ICsupRazao, list(TabelaFinal$Tratamento,
      TabelaFinal$Tempo), function(x) c(x = x))
graficobarrasCaACS1likeRaizIC <- barplot(height = tabelamedias,
      beside = TRUE, las = 1, col=c("darkorchid3", "cyan2", "brown1"),
      ylim = c(0, 3), cex.names = 0.75, ylab = "Razão",
      xlab = "Tempo (horas)", border = "black", axes = TRUE,
      legend.text = TRUE, args.legend = list(title = "Tratamento",
      x = "topleft", bty = "n", cex = .9))
arrows(graficobarrasCaACS1likeRaizIC, tabelaICinf,
      graficobarrasCaACS1likeRaizIC, tabelaICsup, length=0.1, angle=90,
      code=3)

```

**APÊNDICE B – Rotina no programa SAS**

```
/** Lendo o conjunto de dados e criando o fator Trat ***/
data dados;
infile "C:\Users\Renato\Desktop\Andrezza\Dados\CaACS1like\CaACS1like.txt";
input Tecido$ Tratamento$ Tempo          Gene$ Repeticao Replica Cq;
if Tratamento="RW" and Tempo=2 then Trat="RW2 ";
if Tratamento="RW" and Tempo=6 then Trat="RW6 ";
if Tratamento="RW" and Tempo=24 then Trat="RW24 ";
if Tratamento="WD" and Tempo=2 then Trat="WD2 ";
if Tratamento="WD" and Tempo=6 then Trat="WD6 ";
if Tratamento="WD" and Tempo=24 then Trat="WD24 ";
if Tratamento="WW" and Tempo=2 then Trat="WW2 ";
if Tratamento="WW" and Tempo=6 then Trat="WW6 ";
if Tratamento="WW" and Tempo=24 then Trat="WW24 ";
run;

/** Criando o arquivo de dados com a médias das réplicas ***/
proc sort data=dados;
by Tecido Trat Gene Repeticao;
run;

proc means data=dados noprint;
var Cq;
by Tecido Trat Gene Repeticao;
output out=dadosrep mean=Cqmed;
run;

/** Arquivo de dados completo ***/
data Raiz; set dados;
if Tecido="Raiz";
run;

/** Arquivo de dados com média das réplicas*****/
data Raizrep; set dadosrep;
if Tecido="Raiz";
run;
```

```

/*****
/***** ANOVA FIXO *****/
/*****

      /***Ajustando o modelo fixo com a média das réplicas (Cq médio)**/
/*Ajuste do modelo*/
proc mixed plots=none data=Raizrep;
title "Modelo fixo (Cq médio)";
class Trat Gene Repeticao;
model Cqmed = Trat|Gene ;
run; quit;

/*Realizando a transformação Box-Cox (Cálculo de lambda pelo software R)*/
data Raizrep2; set Raizrep;
Cqmedt=log(Cqmed);
run;

/*Ajuste do modelo com Cq médio após transformação Box-Cox*/
proc mixed plots=none data=Raizrep2;
title "Modelo fixo (Cq médio após transformação Box-Cox)";
class Trat Gene Repeticao;
model Cqmedt = Trat|Gene/ solution; *ddfm=sat;
run;

      /*** Ajustando o modelo fixo com todos os valores de Cq
           presentes no conjunto de dados e estrutura de variâncias
           e covariâncias residual ARH(1) ***/
/*Ajuste do modelo*/
proc mixed plots=none data=Raiz;
title "Modelo fixo (todas as réplicas e estrutura da matriz de
      variâncias e covariâncias rsidual ARH(1))";
class Trat Gene Repeticao;
model Cq = Trat|Gene / solution; *ddfm=sat;
repeated / subject=Repeticao type=ARH(1);

```

```

/*Obtendo as estimativas de DDCq e os respectivos EP e IC*/
estimate "RW2-WW2" Trat*Gene 1 -1 0 0 0 0 0 0 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "RW24-WW2" Trat*Gene 0 0 1 -1 0 0 0 0 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "RW6-WW2" Trat*Gene 0 0 0 0 1 -1 0 0 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "WD2-WW2" Trat*Gene 0 0 0 0 0 0 1 -1 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "WD24-WW2" Trat*Gene 0 0 0 0 0 0 0 0 1 -1 0 0 0 -1 1
          0 0 0 0/cl;
estimate "WD6-WW2" Trat*Gene 0 0 0 0 0 0 0 0 0 0 1 -1 -1 1
          0 0 0 0/cl;
estimate "WW24-WW2" Trat*Gene 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 1
          1 -1 0 0/cl;
estimate "WW6-WW2" Trat*Gene 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 1
          0 0 1 -1/cl;

run;

/*****
/***** Modelo Misto (Planta aleatório) *****/
/*****/

/*Ajuste do modelo*/
proc mixed plots=none data=Raiz;
title1 " Modelo Misto (Planta alatorio)";
title2 "Matriz de variâncias ARH(1) - Modela heterogeneidade de variâncias
       de todas as combinacoes de trat e gene e dependencia dos erros";
class Trat Gene Repeticao;
model Cq = Trat|Gene/ solution; *ddfm=sat;
random Repeticao;
repeated / subject=Repeticao type=ARH(1);

```



```
parms / pdata=estparms2;
/*Obtendo as estimativas de DDCq e os respectivos EP e IC*/
estimate "RW2-WW2" Trat*Gene 1 -1 0 0 0 0 0 0 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "RW24-WW2" Trat*Gene 0 0 1 -1 0 0 0 0 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "RW6-WW2" Trat*Gene 0 0 0 0 1 -1 0 0 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "WD2-WW2" Trat*Gene 0 0 0 0 0 0 1 -1 0 0 0 0 0 -1 1
          0 0 0 0/cl;
estimate "WD24-WW2" Trat*Gene 0 0 0 0 0 0 0 0 1 -1 0 0 0 -1 1
          0 0 0 0/cl;
estimate "WD6-WW2" Trat*Gene 0 0 0 0 0 0 0 0 0 0 1 -1 -1 1
          0 0 0 0/cl;
estimate "WW24-WW2" Trat*Gene 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 1
          1 -1 0 0/cl;
estimate "WW6-WW2" Trat*Gene 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 1
          0 0 1 -1/cl;
run;
```