

RAFAEL LÉDO ROCHA DE OLIVEIRA

**IMPUTAÇÃO DE MÉDIAS PARA ANÁLISE DE ESTABILIDADE E  
ADAPTABILIDADE EM EXPERIMENTOS CONJUNTOS  
INCOMPLETOS: UMA APLICAÇÃO EM CAFÉ CONILON**

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

VIÇOSA  
MINAS GERAIS – BRASIL  
2012

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

T

O48i  
2012

Oliveira, Rafael Lédo Rocha de, 1984-

Imputação de médias para análise de estabilidade e adaptabilidade em experimentos conjuntos incompletos: uma aplicação em café conilon / Rafael Lédo Rocha de Oliveira. – Viçosa, MG, 2012.

vii, 52f. : il. (algumas col.) ; 29cm.

Orientador: Cosme Damião Cruz.

Dissertação (mestrado) - Universidade Federal de Viçosa.  
Referências bibliográficas: f. 49-52.

1. Genética - Métodos estatísticos. 2. Melhoramento genético. 3. Biometria. 4. Análise de regressão.  
I. Universidade Federal de Viçosa. II. Título.

CDD 22. ed. 576.5

RAFAEL LÉDO ROCHA DE OLIVEIRA

**IMPUTAÇÃO DE MÉDIAS PARA ANÁLISE DE ESTABILIDADE E  
ADAPTABILIDADE EM EXPERIMENTOS CONJUNTOS  
INCOMPLETOS: UMA APLICAÇÃO EM CAFÉ CONILON**

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

APROVADA: 15 de fevereiro de 2012.

---

Moysés Nascimento

---

Romário Gava Ferrão

---

José Ivo Ribeiro Júnior

---

Cosme Damião Cruz  
(Orientador)

## AGRADECIMENTOS

Agradeço, primeiramente, às pessoas que mais amo: a meu pai João Rocha de Oliveira, exemplo de vida, de caráter, de trabalho, meu ídolo maior; a minha mãe Iracema Lédo do Nascimento Oliveira, exemplo de amor incondicional, de trabalho, de dedicação à família, de bondade e simplicidade; a Ana Lis Lédo Rocha de Oliveira, minha irmã, minha amiga, exemplo de amor à família e humanidade. Essa conquista também é suas!

Ao meu orientador Cosme Damião Cruz, pela paciência, atenção, educação e pelos valiosos conhecimentos transmitidos.

À INCAPER (Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural) e, em especial, ao pesquisador Romário Gava Ferrão pela atenção e fornecimento dos dados experimentais utilizados neste trabalho.

À Universidade Federal de Viçosa, em particular aos professores do DET e DMA, que tanto me ensinaram nestes anos.

À Anita, secretária do DET, pelas brincadeiras e sorriso sempre aberto. Agradeço também ao secretário do PPESTBIO, Joel, pela atenção depositada a todos os assuntos relativos aos alunos do mestrado.

Agradeço também aos amigos que fiz nessa encantadora cidade. Em especial, agradeço a Augusto, Carla, Carlos, Francisco, Daniel, Demétrius, Diogo, Heitor, Isabela, Júlia, Kennedy, Léo, Lucas, Lucius, Mateus, Paulo, Paulo Henrique, Renan, Rodolfo, Rodrigo, Wallace e Vitor.

Agradeço aos meus colegas de mestrado que tanto contribuíram para essa realização. Em especial, a Alyne, Anderson, Camila, Danielle, Diana, Diego, Evandro, Fernanda, Layanne, Lidiane, Liliane, Mariane, Renata, Rodrigo, Suelem e Valéria.

À CAPES e Programa REUNI pela concessão da bolsa de estudos.

## SUMÁRIO

<b>RESUMO</b> .....	<b>iv</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>1. INTRODUÇÃO</b> .....	<b>1</b>
<b>2. REVISÃO BIBLIOGRÁFICA</b> .....	<b>4</b>
2.1. A Imputação de dados .....	4
2.2. Metodologias de estabilidade e adaptabilidade .....	5
2.2.1. Metodologias paramétricas .....	6
2.2.2. Metodologias não-paramétricas .....	12
<b>3. MATERIAL E MÉTODOS</b> .....	<b>14</b>
3.1. Metodologias de Imputação de médias desenvolvidas .....	14
3.1.1. Método I – Modelo com efeito do genótipo .....	14
3.1.2. Método II - Modelo com efeitos do genótipo e ambiente .....	14
3.1.3. Método III - Regressão linear pelos índices ambientais (I) .....	15
3.1.4. Método IV - Regressão linear pelos índices ambientais (II).....	16
3.1.5. Método V - Regressão linear usando genótipos de padrão similar.....	17
3.1.6. Método VI – Manutenção da SQ (GxA).....	19
3.1.7. Ponderações sobre os métodos de imputação.....	21
3.2. Critérios de comparação das metodologias de imputação.....	23
3.2.1. Erro Quadrático Médio .....	24
3.2.2. Coeficiente de correlação de Spearman.....	24
3.2.3. Modificações nas recomendações de Eberhart e Russel .....	25
3.3. O Experimento conjunto completo .....	28
<b>4. RESULTADOS E DISCUSSÃO</b> .....	<b>30</b>
4.1. Análise dos dados .....	30
4.2. Avaliação dos critérios de comparação adotados.....	37
4.2.1. Avaliação do Erro Quadrático Médio .....	37
4.2.2. Avaliação da correlação de Spearman.....	39
4.2.3. Avaliação das modificações nas recomendações de Eberhart-Russel.....	42
<b>5. CONCLUSÕES</b> .....	<b>47</b>
<b>6. REFERÊNCIAS</b> .....	<b>49</b>

## RESUMO

OLIVEIRA, Rafael Léo Rocha de, M.Sc., Universidade Federal de Viçosa, fevereiro de 2012. **Imputação de médias para análise de estabilidade e adaptabilidade em experimentos conjuntos incompletos: uma aplicação em café conilon.** Orientador: Cosme Damião Cruz. Coorientadores: Luiz Alexandre Peternelli, Paulo Roberto Cecon e Leonardo Lopes Bhering.

Este estudo teve por objetivo desenvolver, avaliar e verificar a influência de seis métodos de imputação de médias faltantes na recomendação de genótipos proveniente de metodologias de estabilidade e adaptabilidade em experimentos conjuntos incompletos de café conilon, de modo que seja possível constatar se é conveniente ou não a realização da imputação das médias faltantes. O primeiro método imputa uma média faltante por meio de um modelo que é função da média geral acrescida do efeito do genótipo; já o segundo, adiciona a esse modelo o efeito de ambiente. O terceiro, quarto e quinto métodos elaborados imputam uma média faltante por meio de uma regressão linear simples, cujas variáveis independentes são os índices ambientais, que são estimados utilizando todos os dados disponíveis no experimento conjunto (método 3), ou apenas as médias dos genótipos que foram avaliados em todos os ambientes (método 4). A variável independente do quinto método são as respostas médias do genótipo de maior correlação com aquele que se deseja estimar um valor faltante. A variável dependente destas regressões são as respostas médias do genótipo que se deseja estimar o valor faltante. O sexto método mantém a soma de quadrados da interação genótipo  $\times$  ambiente. Para que a avaliação e verificação da influência dos métodos de imputação propostos fossem possíveis, foi obtido um conjunto de dados cedido pela INCAPER (Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural) oriundo de um experimento conjunto completo em que 38 genótipos (clones) de café conilon foram avaliados em 18 ambientes segundo suas produtividades (sacas/hectare). Este experimento foi submetido às análises de estabilidade e adaptabilidade de Lin & Binns (1998) e Eberhart-Russel (1966), gerando, dessa forma, recomendações padrão. Feito isto, médias foram retiradas aleatoriamente simulando experimentos com 1%, 5% e 10% de médias faltantes. Daí, por meio dos métodos de imputação elaborados, novos valores de médias foram gerados. A avaliação dessas metodologias e a verificação da influência da imputação na recomendação dos genótipos foram realizadas por meio do cálculo

do Erro Quadrático Médio, da Correlação de Spearman entre a recomendação de Lin & Binns antes e após a imputação das médias, e da porcentagem de mudanças na recomendação dos genótipos em relação à recomendação padrão de Eberhart-Russel. Conforme os resultados obtidos neste estudo, a realização da imputação das médias mediante as metodologias desenvolvidas com melhor desempenho (2, 3, 4 e 6) em experimentos conjuntos incompletos é aconselhável, uma vez que a alteração nas recomendações dos genótipos avaliados foi pequena se comparado com o número de médias faltantes nos ensaios avaliados.

## ABSTRACT

OLIVEIRA, Rafael Léo Rocha de, M.Sc., Universidade Federal de Viçosa, February, 2012. **Imputation of data for the stability and adaptability study in incomplete experiments sets: an application to data of conilon coffee.** Advisor: Cosme Damião Cruz. Co-advisors: Luiz Alexandre Peternelli, Paulo Roberto Cecon and Leonardo Lopes Bhering.

This study aimed to develop, evaluate and verify the influence of medium values imputation through six methods on the genotype recommendation imputing missing in medium genotype from methodologies of stability and adaptability in experiments incomplete sets of conilon coffee, so that it can determine whether it is appropriate or not the implementation of missing means imputation. The first method imputes a missing value through a mathematical model that is a function of the experimental medium value added the effect of genotype; the second method adds to this model the environmental effect. The third, fourth and fifth methods created impute a missing value through a simple linear regression, whose independent variables are the environmental effects, which are estimated using all available data in the experiment set (method 3), or just the means of the genotypes that were evaluated in all environments (method 4). The independent variable of the fifth method are the mean responses of the genotype most correlated with the one you want to estimate a missing value. The dependent variable of these regressions are the mean responses of the genotype that you want to estimate the missing value. The sixth method keeps the sum of squares of the genotype x environment interaction. In order to evaluate and verify the influence of imputation methods proposed, it was obtained a data set provided by INCAPER (Capixaba Institute of Research, Technical Assistance and Rural Extension). It was from an experiment tin which 38 complete set genotypes (clones) of conilon coffee were evaluated in 18 environments according to your productivity (bags/hectare). This experiment was submitted to Lin & Binns (1998) and Eberhart-Russell (1966) stability and adaptability analyses, generating, then, reference recommendations. After that, some means were removed randomly simulating experiments with 1%, 5% and 10% of missing means. Then, new mean values were generated through the imputation methods developed. The evaluation of these methodologies and verification of the influence of imputation on the recommendation of the genotypes was performed by calculating the mean squared



error, the Spearman correlation between the Lin & Binns recommendation before and after imputation, and the percentage of changes on the recommendation of the genotypes in relation to the reference Eberhart-Russel recommendation. According to the results obtained in this study, the performance of the imputation of means by the developed methodologies with better performance (2, 3, 4 and 6) in incomplete experiment sets is advisable, since the changes in the recommendations of the genotypes was small compared with the number of missing means in the evaluated experiments.

## 1. INTRODUÇÃO

O principal objetivo dos programas de melhoramento de plantas é o aumento da produtividade de grãos ou frutos. Geralmente, um novo genótipo só é lançado no mercado quando ele apresenta melhor desempenho e adaptabilidade mais ampla que os genótipos que estão sendo plantados pelo agricultor (BESPALHOK, OLIVEIRA e GUERRA, 2008). Obviamente, há também preocupação com a melhoria na qualidade, da resistência a pragas e doenças, facilidade de cultivo e menor exigência nutricional, dentre outras características.

Em experimentos conjuntos, diferentes genótipos são avaliados quanto a sua produtividade em diversos ambientes (locais, épocas, anos). A última etapa em um programa de melhoramento consiste na recomendação desses genótipos. Porém, como relatam Cruz e Regazzi (1997), quando se consideram uma série de ambientes, detectam-se, além dos efeitos genéticos e ambientais, um efeito adicional, proporcionado pela interação entre os mesmos.

A interação genótipos x ambientes ocorre pela reposta diferenciada dos materiais genéticos nos diferentes ambientes, porém, Eberhart e Russel (1996) afirmam que esse fato faz parte da evolução das espécies. A ausência da referida interação, segundo Kang (1998), pode indicar insuficiência de diversidade genética, o que seria desastroso por causa da vulnerabilidade genética da espécie associada a doenças epidêmicas, infestação de insetos ou fatores bióticos.

A interação genótipos x ambientes classifica-se em dois tipos: simples e complexa. A primeira é proporcionada pela diferença de variabilidade entre os genótipos nos ambientes, de forma que a posição relativa dos genótipos geralmente não é alterada. A segunda ocorre pela falta de correlação entre os desempenhos dos genótipos, de modo que estes apresentam diferentes respostas às variações ambientais, causando, frequentemente, alteração na sua classificação, considerando os diversos ambientes (CRUZ e CARNEIRO, 2003).

Cruz e Regazzi (1997) afirmam que a avaliação da interação genótipos x ambientes torna-se de grande importância para o melhoramento, pois, no caso de sua existência, há possibilidades de o melhor genótipo em um ambiente não o ser em outro. Assim, cabe ao melhorista avaliar sua magnitude e significância e

fornecer subsídios que possibilitem adotar procedimentos para sua minimização e, ou, seu aproveitamento.

Estudos acerca de tal interação, apesar de ser de grande importância para o melhoramento, não proporcionam informações pormenorizadas sobre o comportamento de cada genótipo frente às variações ambientais (CRUZ e REGAZZI, 1997). Conforme afirmam Schmildt e Cruz (2005), duas alternativas práticas podem ser realizadas com o objetivo de minimizar os efeitos da interação genótipos x ambientes. A primeira delas é a estratificação ambiental, permitindo que sejam formados estratos que contenham sub-regiões mais homogêneas, o que poderá dar subsídios ao descarte de ambientes em caso de escassez de recursos. A segunda alternativa, que particulariza os genótipos, é o emprego daqueles com ampla adaptabilidade e boa estabilidade.

Estudos de estabilidade e adaptabilidade tornam possível a identificação de genótipos de comportamento previsível e que sejam responsivos a variações ambientais em condições específicas ou amplas (CRUZ e REGAZZI, 1997). Assim, a análise de estabilidade e adaptabilidade auxilia na tomada de decisões de um programa de melhoramento genético no momento da recomendação dos genótipos (FALCONER e MACKAY, 1998).

Métodos estatísticos têm sido propostos e utilizados em diferentes situações, e, com frequência, novas metodologias são utilizadas para interpretar melhor a interação genótipo x ambiente (PELÚZIO *et al.*, 2008). Segundo Cruz e Regazzi (1997), a escolha de um método de análise depende dos dados experimentais, principalmente os relacionados com o número de ambientes disponíveis, da precisão requerida e do tipo de informação desejada. Os autores ainda salientam que alguns métodos são alternativos, outros são complementares, podendo ser utilizados conjuntamente. No entanto, em experimentos conjuntos incompletos, o uso das metodologias de estabilidade e adaptabilidade é comprometido.

A incompletude mencionada não se caracteriza pela perda de parcela(s) de um determinado genótipo em certo bloco. O interesse aqui se recai na situação em que um material genético não é avaliado em certo ambiente (ataque de pragas ou doenças, inviabilidade das sementes, insuficiente quantidade de mudas para todos os ambientes, problemas de mão-de-obra), fato este bastante comum no cotidiano de um experimento.

A literatura registra várias metodologias de imputação de médias faltantes, entretanto não foram encontrados trabalhos que apontassem a influência da aplicação dessas metodologias nas recomendações dos genótipos provenientes de um estudo de estabilidade e adaptabilidade; ou seja, não foram encontrados estudos que verificassem se é razoável imputar médias faltantes no experimento conjunto de modo que não ocorram mudanças significativas nos resultados das análises de estabilidade e adaptabilidade dos genótipos em avaliação.

Diante do exposto, este trabalho visou desenvolver e avaliar novos métodos de imputação de médias faltantes em experimentos conjuntos incompletos para fins de estudo de estabilidade e adaptabilidade em café conilon.

Além disso, o estudo teve como objetivo verificar a influência da imputação das médias não avaliadas na recomendação dos genótipos provenientes das análises de estabilidade e adaptabilidade, de modo que se possa constatar se é conveniente ou não a realização da imputação das médias faltantes. Espera-se que este trabalho ofereça contribuições a pesquisadores que se encontrem em situações semelhantes às aqui avaliadas, e, dessa forma, o estudo os auxiliem a tomar a melhor decisão (imputar médias faltantes segundo uma das metodologias aqui expostas ou seguir com o experimento incompleto).

Deve ser ressaltado que as metodologias desenvolvidas neste trabalho não são exclusivas do café conilon. Portanto, tais métodos podem ser utilizados em qualquer cultura.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1 A Imputação de dados

Como relata Arciniegas-Alarcón e Dias (2009), muitas vezes, em experimentos multiambientais, muitos genótipos não são testados em alguns ambientes; entretanto, pode ser de interesse do pesquisador obter estimativas do desempenho de combinações que não foram testadas, e tais estimativas podem ser calculadas utilizando informações daquelas combinações genótipo por ambiente (G x A) que foram observadas.

Arciniegas-Alarcón et al. (2011) afirmam que os primeiros estudos sobre imputação de dados (médias) faltantes foram realizados por Freeman (1975), em que ele gerou valores de modo iterativo minimizando a soma de quadrados do erro. Alternativas para a imputação de dados (médias) faltantes também foram sugeridas por Gauch e Zobel (1990), que fizeram a imputação através do uso do algoritmo EM e o modelo AMMI (modelos de efeitos aditivos de interação multiplicativa), mas algumas alternativas desse procedimento usando estatística multivariada (análises de agrupamento) foram descritas em Godfrey *et al.* (2002). Mandel (1993) propôs fazer a imputação em tabelas incompletas de dupla entrada usando funções lineares das linhas (ou colunas).

Estudos com este mesmo fim foram desenvolvidos por Denis e Baril (1992) e Calinskiet al. (1992). Conforme Arciniegas-Alarcón e Dias (2009) relatam, eles encontraram que imputações por meio de modelos AMMI usando submodelos robustos ou baseados em mínimos quadrados alternados atingem resultados tão bons como os encontrados com um algoritmo EM. Bergamo, Dias e Krzanowski (2008) propuseram um método baseado em imputação múltipla livre de distribuição, o qual pode ser aplicado em matrizes de interação G x A com informação incompleta.

## 2.2 Metodologias de estabilidade e adaptabilidade

Segundo Molina (2007), diversos métodos têm sido propostos para a avaliação da estabilidade e adaptabilidade fenotípica, e a diferença entre esses métodos é baseada nos próprios conceitos e procedimentos biométricos para quantificar a interação. Como afirmam Mariotti *et al.* (1976), há dificuldades para o esclarecimento dos termos estabilidade e adaptabilidade devido às várias definições sugeridas por diversos autores.

Conforme Lavoranti (2003) relata, os principais conceitos para o termo estabilidade são os seguintes: capacidade de um indivíduo ou de uma população de produzir variações mínimas de fenótipos em diferentes ambientes; potencial para ajustar o seu estado genotípico e fenotípico às flutuações ambientais; capacidade de os genótipos mostrarem um comportamento altamente previsível em função do estímulo ambiental. Ainda segundo Lavoranti (2003), outros autores conceituam estabilidade da seguinte maneira: capacidade do genótipo de obter respostas aos ambientes que sejam paralelas ao desempenho médio de todos os genótipos avaliados nos experimentos; consistência no ranqueamento relativo de outros genótipos, dentre outros. A adaptabilidade refere-se à capacidade de os genótipos aproveitarem vantajosamente os estímulos do ambiente

Becker & Leon (1988) dividem estabilidade em dois tipos: estática, associando aqueles genótipos que apresentam desempenho constante frente às variações ambientais. Esse tipo de estabilidade é desejável quando se quer preservar determinada característica genética; e dinâmico, associando os genótipos que apresentam um comportamento previsível dentro das variações ambientais.

É importante ressaltar que os parâmetros que determinam a estabilidade são específicos para os materiais genéticos avaliados, bem como para os ambientes analisados, sendo inadequado, portanto, inferir esses resultados para outros materiais e ambientes (YATES e COCHRAN, 1938).

Dentre as metodologias paramétricas univariadas de estabilidade e adaptabilidade sugeridas pelas literaturas, destacam-se aquelas baseadas em análise de variância, regressão linear simples, regressão linear bi-segmentada, regressão quadrática e análises não-lineares. Também há metodologias não-paramétricas e fundamentadas em análises multivariadas.

### 2.2.1 Metodologias paramétricas

A primeira metodologia que teve o objetivo de avaliar a estabilidade de genótipos foi a tradicional. Nela, por meio da análise do experimento conjunto, a estimativa de estabilidade é feita por meio da variação da resposta de cada genótipo nos ambientes avaliados; aquele genótipo que obtiver menor quadrado médio é o mais estável. Tal conceito de estabilidade, segundo Cruz e Regazzi (1997), é o principal defeito deste método, uma vez que ele vem de encontro aos propósitos de melhoramento, pois os genótipos que apresentam menores variâncias entre ambientes são, em geral, os menos produtivos.

Os métodos de Plaisted e Peterson (1959) e de Wricke (1962) são também métodos que quantificam somente a estabilidade baseados na análise de variância. Segundo Rocha *et al.* (2009), o método de Wricke tem sido utilizado para avaliar a estabilidade fenotípica em soja; este método utiliza a variância da interação genótipos x ambientes e realiza o desdobramento em componentes atribuídos a cada genótipo. Essa metodologia, conforme afirmam Franceschi *et al.* (2010), recomenda genótipos estáveis independente do rendimento médio e de suas respostas a ambientes favoráveis e desfavoráveis, não sendo conveniente, portanto, sua utilização quando o objetivo é recomendar genótipos a condições específicas de ambiente.

Conforme relatam Cruz e Regazzi (1997), boas concordâncias entre os métodos de Plaisted e Peterson (1959) e de Wricke (1962) já foram constatadas experimentalmente. Outros métodos de estabilidade baseados em análises de variância são os de Shukla (1972), Annicchiarico (1992) e Magari e Kang (1997).

Métodos como os de Finlay e Wilkinson (1963) e de Eberhart e Russel (1966) baseiam-se em análises de regressão linear que medem o desempenho de cada genótipo frente às mudanças ambientais.

Para os dois métodos citados, é adotado o seguinte modelo de regressão linear:

$$Y_{ij} = \beta_{0i} + \beta_{1i} \cdot I_j + \delta_{ij} + \bar{\epsilon}_{ij}$$

Em que:

$Y_{ij}$  = média do genótipo  $i$  no ambiente  $j$ ;

$\beta_{0i}$  = constante da regressão;

$\beta_{1i}$  = coeficiente de regressão;

$\delta_{ij}$  = desvio da regressão;

$\bar{\epsilon}_{ij}$  = erro experimental médio;

$I_j$  = índice ambiental codificado pela média que é dada pela seguinte expressão:

$$I_j = \bar{Y}_j - \bar{Y}_{..}$$

O índice ambiental do local  $I_j$  codificado pela média é dado pela média dos genótipos avaliados no ambiente  $j$  subtraída da média geral. Assim, um índice ambiental positivo indica um ambiente favorável, ao passo que um índice ambiental negativo indica um ambiente desfavorável. Para as análises por meio do método proposto por Finlay e Wilkinson (1963), os dados são transformados para a escala logarítmica com o intuito de que seja induzido um grande grau de linearização; o que não se faz necessário na metodologia de Eberhart e Russel (1966). Então, para cada genótipo, é estimada uma regressão linear simples que nos fornece a resposta do genótipo  $i$  ambiente  $j$ .

Tanto para os métodos de Finlay e Wilkinson (1963), quanto para o de Eberhart e Russel (1966), o coeficiente de regressão igual a 1 indica que o genótipo apresenta adaptabilidade a toda rede experimental; valores menores que 1 indicam uma adaptabilidade a ambientes desfavoráveis; ao passo que valores maiores que 1 para tal coeficiente denotam adaptabilidade a ambientes favoráveis. O conceito de adaptabilidade nas duas metodologias expressa, então, a resposta do genótipo às variações ambientais; a adaptabilidade seria a capacidade de o genótipo responder vantajosamente à mudança ambiental.

Para o método de Finlay e Wilkinson (1963), a estabilidade se dá pela invariância de respostas avaliadas em diferentes locais. Assim, o coeficiente de regressão passa a ser parâmetro único para os dois conceitos. Por exemplo, uma



estimativa para o coeficiente de regressão  $\beta_{1i}$  maior que 1 indica, como mencionado, adaptabilidade específica para ambientes favoráveis e, no caso, estabilidade baixa. Segundo Cruz e Regazzi (1997), são também considerados de alto risco para o cultivo em regiões de baixa tecnologia, pois reduzem significativamente seu desempenho em ambientes desfavoráveis. Segundo o método de Eberhart e Russel (1966), a magnitude e a significância da variância dos desvios da regressão dão uma idéia da previsibilidade do material estudado, que é exatamente o que tais autores passaram a definir como estabilidade. Ou seja, serão considerados genótipos estáveis aqueles que obtiverem valores pequenos para o desvio da regressão.

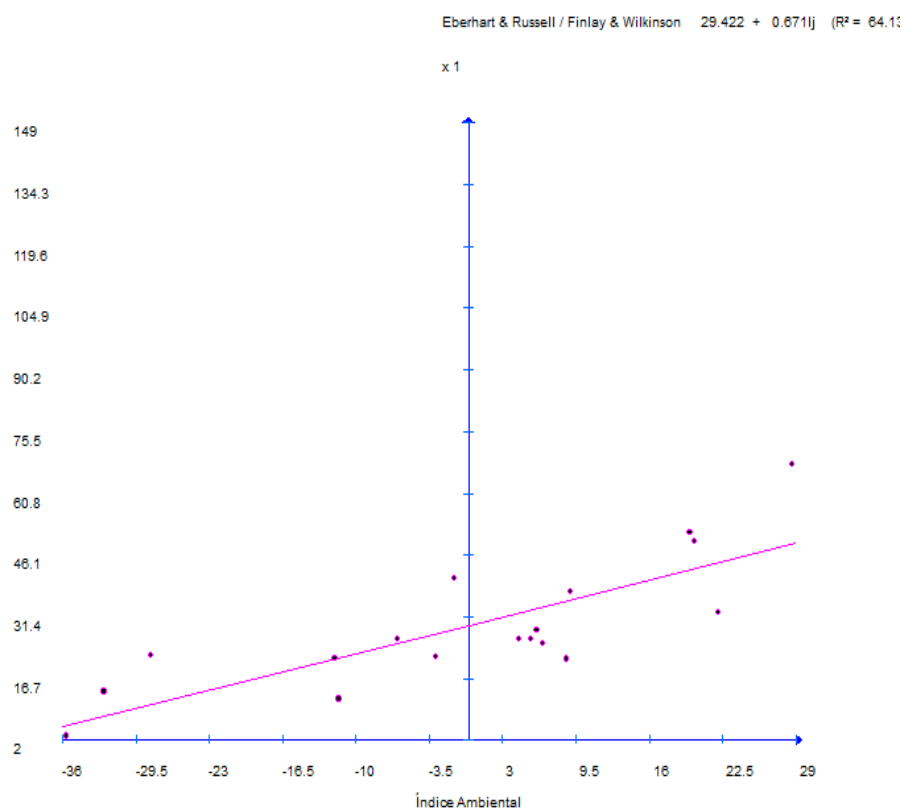
Ainda conforme Cruz e Regazzi (1997), o método de Eberhart e Russel (1966) define estabilidade como a capacidade de os genótipos mostrarem um comportamento altamente previsível em função do estímulo do ambiente, e tal conceito é avaliado pelo componente de variância atribuído aos desvios da regressão ( $\sigma^2_{di}$ ), sendo verificados os seguintes tipos de genótipos:

- Genótipos com estabilidade ou previsibilidade alta: são aqueles com  $\sigma^2_{di}$  igual a 0;
- Genótipos com estabilidade ou previsibilidade baixa: são aqueles com  $\sigma^2_{di}$  maior que 0.

É evidente que a recomendação de determinado genótipo para um ambiente ou grupo de ambientes quaisquer é sempre feita quando tal genótipo possui alta produtividade média relativa. Assim, Cruz e Regazzi (1997) relatam que Eberhart e Russel (1966) consideram como genótipo ideal aquele que apresenta alta produtividade média, coeficiente de regressão igual a 1,0 e desvios de regressão tão pequenos quanto possíveis. Analogamente, o genótipo ideal para ambientes favoráveis é aquele com alta produção média, coeficiente de regressão maior que 1,0 e desvios de regressão tão pequenos quanto possíveis; ao passo que o genótipo a ser recomendado para ambientes desfavoráveis possui alta produção média, coeficiente de regressão menor que 1,0 e desvios de regressão tão pequenos quanto possíveis.

É importante explicitar que a recomendação de genótipos para ambientes desfavoráveis não deve ser considerada de pouca importância, uma vez que é por meio dela que pequenos produtores, proprietários de terras que utilizam menos tecnologias podem se beneficiar.

Visando um melhor entendimento, as Figuras 1, 2 e 3 mostram o comportamento de três genótipos em vários ambientes e suas respectivas equações de regressão linear estimadas pela metodologia de Eberhart e Russel (1966). A variável presente no eixo das ordenadas nas três figuras é a produtividade (sacas beneficiadas/hectare).

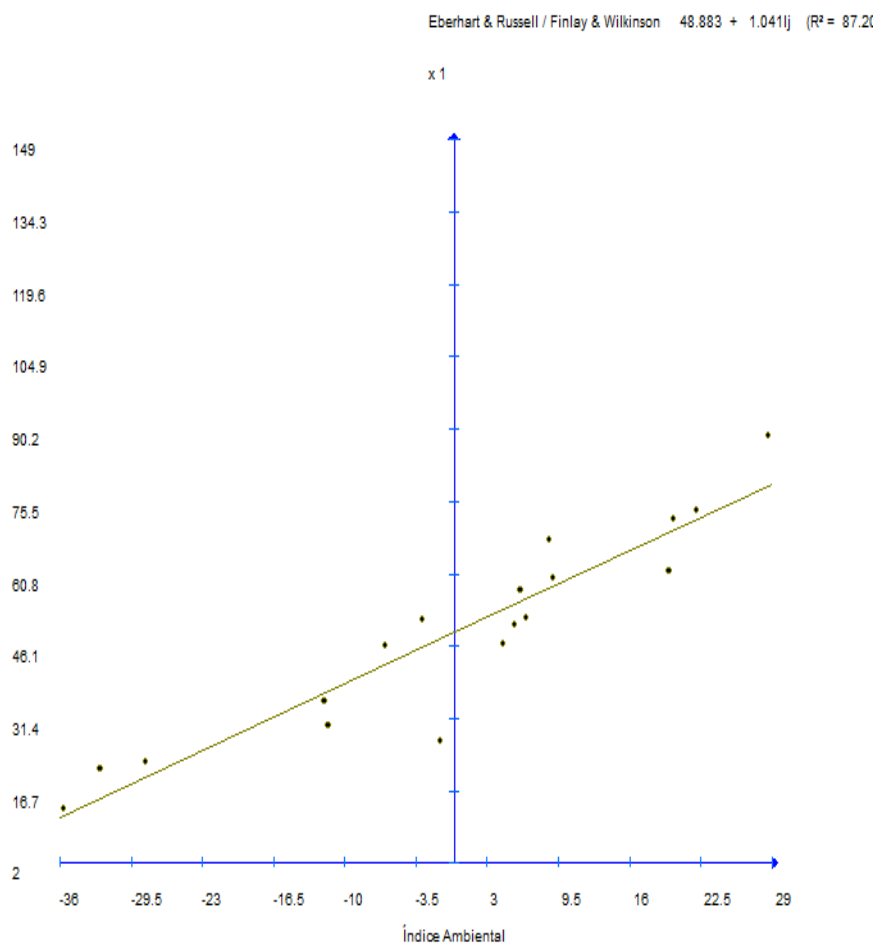


**Figura1.** Comportamento do genótipo com menor estimativa do parâmetro de adaptabilidade (GENES, 2006).

Na parte superior do gráfico da Figura 1, nota-se a equação de regressão linear estimada para o genótipo em questão pelo método de Eberhart e Russel (1966). Verifica-se que o coeficiente de regressão linear estimada é menor que 1, sendo este então um indicativo para sua recomendação a ambientes desfavoráveis. Entretanto, o experimento retratado na Figura 1 possui média geral igual a 48,75;

que é maior que a média deste genótipo, a qual pode ser verificada pela estimativa da constante de regressão (29,422).

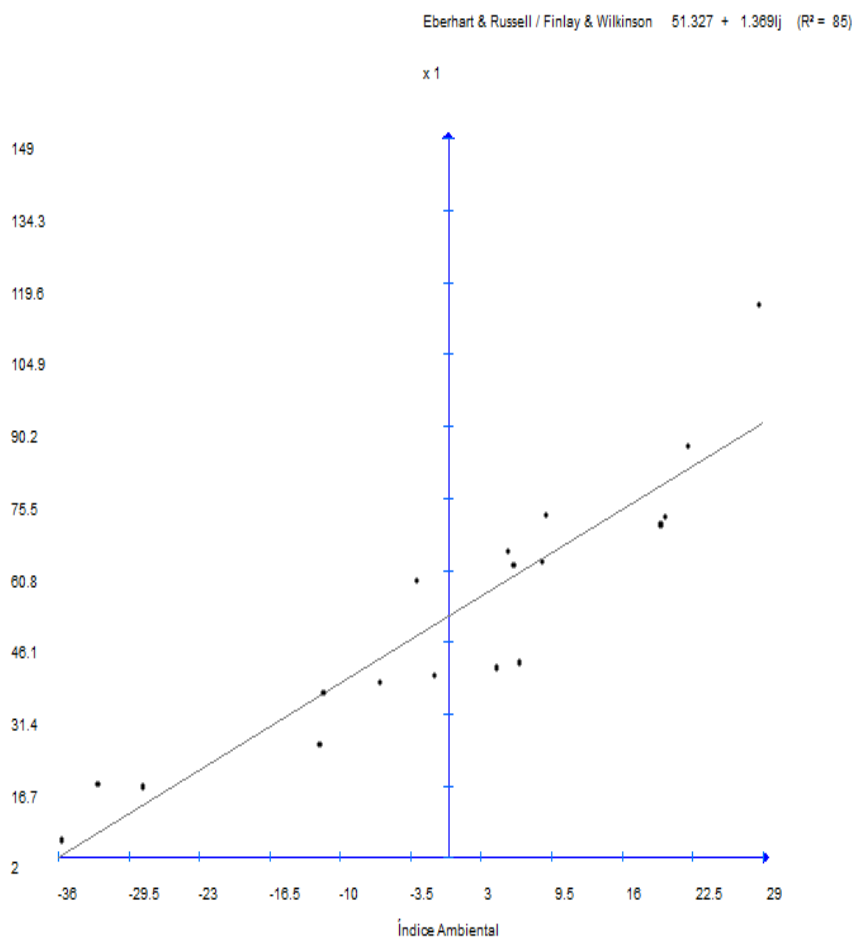
É importante também levar em consideração o coeficiente de determinação ( $R^2$ ), que pode ser uma medida auxiliar para a tomada de decisão na recomendação de um genótipo.



**Figura2.** Comportamento do genótipo com estimativa do parâmetro de adaptabilidade mais próximo de 1 (GENES, 2006).

Para este segundo exemplo, foi estimado, por meio da metodologia de Eberhart e Russel, um valor para o coeficiente de regressão mais próximo de 1, o que o torna mais adequado a ser recomendado para toda a rede experimental. Ainda, a média é superior à média geral do experimento e o seu valor do coeficiente de determinação é relativamente alto, o que são indícios de alta previsibilidade.

Obviamente, as conclusões acerca de sua adaptabilidade geral e alta previsibilidade devem ser tomadas diante do resultado de um teste estatístico.



**Figura3.** Comportamento do genótipo com a maior estimativa do parâmetro de adaptabilidade (GENES, 2006).

Neste terceiro e último exemplo, nota-se uma estimativa para o coeficiente de regressão maior que 1, uma média maior que a média geral do experimento e um valor relativamente alto para  $R^2$ . Além disso, constata-se pela análise do gráfico que o genótipo não possui bom desempenho em ambientes desfavoráveis, diminuindo bastante seu rendimento com o decréscimo do índice do ambiente, fato este justificado pelo alto valor de seu coeficiente de regressão estimado.

De acordo com Molina (2007), outras metodologias baseadas em regressão linear simples são as de Perkins e Jinks (1968) e Tai (1971).

Conforme relata Lavoranti (2003), Verma et al. (1978) propuseram uma análise em que fossem estimados duas regressões lineares para cada genótipo; a

primeira considerando os ambientes desfavoráveis, e a segunda considerando os ambientes favoráveis incluindo também o ambiente com menor índice negativo em módulo. Tal metodologia foi motivada pelo fato de um genótipo ideal apresentar alta capacidade para responder a ambientes favoráveis e invariante nos ambientes desfavoráveis. É notado que nas metodologias anteriores, é impossível de se identificar tal genótipo ideal, uma vez que se avalia uma única análise de regressão.

É importante salientar que, como são estimadas duas regressões lineares por genótipo, seriam necessários, no mínimo, seis ambientes no experimento. Conforme Cruz e Carneiro (2003), Silva e Barreto (1985) propuseram o ajustamento de apenas uma única equação de regressão constituída de dois segmentos de reta, com união no ponto correspondente ao valor zero do índice de ambiente. Cruz, Torres e Vencovsky (1989) apresentaram extensão da metodologia proposta por Silva e Barreto, tornando-a operacionalmente mais simples e com propriedades estatísticas mais adequadas aos propósitos do melhoramento. Como cita Molina (2007), outra metodologia baseada em regressão linear bi-segmentada é a de Storck e Vencovsky (1994).

Como exemplo de metodologia baseada em regressão quadrática, tem-se o método de Brasil e Chaves (1994). Baseadas em análises não lineares, há as metodologias de Chaves, Vencovsky e Geraldi (1989); Toller e Burrows (1998); Silva (1998) e Rosse e Vencovsky (2000). Há, ainda, o método de Cornelius, Crossa e Seyedsadr (1996) que é baseado em análises multivariadas que integram a análise comum de variância com a análise de componentes principais (MOLINA, 2007).

### **2.2.2 Metodologias não-paramétricas**

Podem-se citar metodologias de adaptabilidade e estabilidade não-paramétricas, a saber: Lin & Binns (1988), Huehn (1990), Fox *et al.* (1990), Carneiro (1998), dentre outros. De acordo com Cruz e Carneiro (2003), tais metodologias apresentam as seguintes vantagens em relação aos métodos paramétricos, dentre outras: não é necessário assumir qualquer hipótese sobre a distribuição dos valores fenotípicos; as medidas estimadas com base na

classificação são de fácil uso e interpretação; a retirada ou adição de um ou poucos genótipos não seria causa de grandes variações nas estimativas.

A metodologia de Lin & Binns (1998) possui como estimativa de estabilidade e adaptabilidade de genótipos o quadrado médio da distância entre a média do genótipo e a resposta média máxima obtida no ambiente (FRANCESCHI *et al.*, 2010). Tal estimativa é dada pela seguinte expressão:

$$P_i = \frac{\sum_{j=1}^n (Y_{ij} - M_j)^2}{2n}$$

Em que:

$P_i$  = estimativa da estabilidade e adaptabilidade do  $i$ -ésimo genótipo;

$Y_{ij}$  = produtividade do  $i$ -ésimo genótipo no  $j$ -ésimo ambiente;

$M_j$  = resposta máxima observada entre todos os genótipos no  $j$ -ésimo ambiente;

$n$  = número de ambientes.

O menor valor para a estatística  $P_i$  indica o genótipo a ser recomendado para toda rede experimental. Cruz e Carneiro (2003) relatam que a regressão do genótipo dito ideal (resposta máxima em todos os ambientes) apresentará coeficiente de regressão mais próximo da unidade. Assim, o genótipo que apresentar menor valor para  $P_i$  é aquele que mais se aproxima da estimativa do coeficiente de regressão igual a 1. Ainda, é razoável que se conclua que o genótipo com menor valor de  $P_i$  é também aquele com menores flutuações de desempenho frente às variações ambientais dentre os genótipos com melhor rendimento. Desse modo, considerando o conceito de estabilidade fenotípica, o genótipo com menor valor para  $P_i$  tem adaptabilidade geral, alta estabilidade e alto rendimento; sendo assim recomendável para toda rede experimental.

### 3 MATERIAL E MÉTODOS

#### 3.1 Metodologias de Imputação de médias desenvolvidas

Este estudo tem como objetivo desenvolver e avaliar métodos de imputação de médias faltantes em experimentos conjuntos incompletos para fins de estudo de estabilidade e adaptabilidade. Foram elaboradas seis metodologias de imputação de médias faltantes, a saber:

##### 3.1.1 Método I – Modelo com efeito do genótipo

Essa metodologia se baseia em imputar a média de um genótipo em determinado ambiente por meio de um modelo aditivo que seja função da média geral experimental e do efeito de tal genótipo no ensaio.

O modelo adotado é:

$$Y_{ij} = \hat{\mu} + \hat{G}_i$$

Em que:

$Y_{ij}$  = média imputada do genótipo  $i$  no ambiente  $j$ ;

$\hat{\mu}$  = média geral do experimento;

$\hat{G}_i$  = efeito do genótipo  $i$ .

##### 3.1.2 Método II - Modelo com efeitos do genótipo e ambiente

Essa metodologia se baseia em imputar o valor de média de um genótipo em determinado ambiente por meio de um modelo aditivo que seja função da média geral experimental, do efeito de tal genótipo no ensaio e do efeito do ambiente em que não houve a avaliação da referido genótipo.

O modelo adotado é:

$$Y_{ij} = \hat{\mu} + \hat{G}_i + \hat{A}_j$$

Em que:

$Y_{ij}$  = média imputada do genótipo  $i$  no ambiente  $j$ ;

$\hat{\mu}$  = média geral do experimento;

$\hat{G}_i$  = efeito do genótipo  $i$ ;

$\hat{A}_j$  = efeito do ambiente  $j$ .

### 3.1.3 Método III - Regressão linear pelos índices ambientais (I)

Este método consiste na obtenção de uma equação de regressão linear simples que mede a resposta de um determinado genótipo em certo ambiente em função do seu índice ambiental.

Como o interesse aqui se encontra na estimação das médias não avaliadas, estimam-se, então, equações de regressão linear para aqueles genótipos em que não foi possível a avaliação de, ao menos, uma média. Nesta metodologia, cada índice ambiental é calculado levando em consideração todos os genótipos avaliados naquele local.

Após a obtenção dos índices ambientais, para cada genótipo em que há pelo menos um valor faltante, é estimada uma equação de regressão linear mediante  $(a-f)$  pontos, em que  $a$  representa o número de ambientes no experimento, e  $f$  indica o número de médias faltantes do genótipo cuja equação está sendo estimada.

As abscissas desses pontos são os índices ambientais cujos ambientes foram avaliados pelo genótipo em questão, e as ordenadas são as respostas desse genótipo nesses ambientes. Dessa maneira, é estimada uma equação de regressão linear para tal genótipo. Para estimar suas médias faltantes, basta que se atribua à variável independente da equação estimada, os valores dos índices cujos ambientes não puderam ser avaliados pelo referido genótipo. Os valores da variável dependente resultantes dessas atribuições serão, então, as médias estimadas desejadas.



Dessa maneira, como na metodologia de Eberhart e Russel, o modelo estatístico adotado é:

$$Y_{ij} = \beta_0 + \beta_1 I_j + \delta_{ij} + \bar{\varepsilon}_{ij}$$

Em que:

$$I_j = \bar{Y}_j - \bar{Y}_{..}$$

Neste método, tem-se que:

$$\bar{Y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} \text{ e } \bar{Y}_{..} = \frac{\sum_i \sum_j Y_{ij}}{\sum_{j=1}^a n_j}; \text{ em que:}$$

$n_j$  = número de genótipos avaliados no ambiente  $j$ .

$\bar{\varepsilon}_{ij}$  = erro aleatório médio.

### 3.1.4 Método IV - Regressão linear pelos índices ambientais (II)

Esta metodologia de estimação é bastante similar à anterior. Elas se diferenciam basicamente pelo modo através do qual os índices ambientais são estimados. Nesta metodologia, a estimação do índice ambiental é dada pela média dos genótipos completos (genótipos avaliados em todos os ambientes) observados em cada ambiente subtraída da média geral dos valores avaliados envolvendo apenas os genótipos completos em todo o experimento.

Assim, o modelo estatístico adotado é:

$$Y_{ij} = \beta_0 + \beta_1 I_j + \delta_{ij} + \bar{\varepsilon}_{ij}$$

Em que:

$$I_j = \bar{Y}_j - \bar{Y}_{..}$$

Considera-se, para o cálculo de  $I_j$ , apenas o conjunto de  $c$  genótipos que foram avaliados em todos os ambientes, ou seja:

$$\bar{Y}_{.j} = \frac{1}{c} \sum_{i=1}^c Y_{ij} \text{ e } \bar{Y}_{..} = \frac{1}{c \cdot a} \sum_{i,j=1}^{c,a} Y_{ij}$$

### 3.1.5 Método V - Regressão linear usando genótipos de padrão similar

Esta metodologia consiste na obtenção das estimativas dos coeficientes de correlação de Pearson entre os desempenhos médios de pares de genótipos existentes no experimento. Os cálculos de correlação entre dois genótipos quaisquer são realizados levando em consideração os ambientes em que houve a avaliação dos dois genótipos entre os quais se deseja estimar sua correlação.

Assim, caso haja no experimento um valor faltante para algum genótipo, verifica-se qual é o maior valor de correlação que envolve tal genótipo e os demais. Daí, a equação de regressão linear que será estimada para o genótipo incompleto citado é baseada apenas nos dados desses dois genótipos envolvidos na maior correlação calculada, e não mais são levados em consideração os dados restantes; como ocorre com o uso das metodologias III e IV.

Muitas vezes, porém, seguindo o algoritmo descrito, os coeficientes de correlação de Pearson entre os desempenhos médios de pares de genótipos seriam calculados mediante um diferente número de pares de pontos. Isso é possível, pois o número de ambientes no qual um determinado genótipo, por ventura, não pôde ser avaliado nem sempre é igual para todos os genótipos. Entretanto, é sabido que a estimativa do valor do coeficiente de correlação de Pearson ( $r$ ) tende a diminuir quando aumenta o número de pares de pontos fornecidos para gerar uma equação de regressão linear simples; logo, o cálculo de tal coeficiente de correlação não é apropriado nestas situações.

Uma alternativa para a solução deste problema é o cálculo do coeficiente de determinação ajustado ( $R^2_{aj.}$ ) para cada par de genótipos, que é expresso pela seguinte expressão:

$$R^2_{aj.} = 1 - \left[ \frac{(N - 1)(1 - R^2)}{N - K} \right]$$

Em que:

$R^2$  = coeficiente de determinação;

$N$  = número de pares de pontos;

$K$  = número de parâmetros do modelo de regressão utilizado.

Nota-se que, em um modelo de regressão linear simples, o valor de  $K$  é dois; assim, com exceção do caso em que o coeficiente de determinação seja igual a um, o valor do coeficiente de determinação ajustado será sempre menor que  $R^2$ . Ainda, verifica-se que mantendo o valor de  $R^2$  e aumentando o valor de  $N$ , o valor entre colchetes diminui, aumentando o valor do  $R^2_{aj.}$ ; o que é razoável, visto que, como mencionado anteriormente, maiores valores de  $N$ , tendem a diminuir o valor da estimativa do coeficiente de correlação de Pearson.

De modo análogo, é detectada a maior estimativa do coeficiente de determinação ajustado que envolve o genótipo com média(s) faltante(s) e, assim, é estimada uma equação de regressão linear que é baseada apenas nos dados dos dois genótipos envolvidos na maior estimativa de coeficiente de determinação ajustado calculada ( $R^2_{aj.}$ ).

A estimação da equação de regressão linear simples é realizada por meio de  $a_g$  pontos, em que  $a_g$  denota o número de ambientes em que houve a avaliação dos dois genótipos envolvidos na regressão. As abscissas desses pontos são as médias avaliadas nestes  $a_g$  ambientes do genótipo que contém a maior correlação/coeficiente de determinação ajustado com o genótipo não-completo. As ordenadas são as respostas do genótipo não-completo nestes mesmos ambientes. Daí, estima-se uma equação de regressão linear simples mediante tais pontos. Em seguida, basta que se atribua à variável independente dessa equação, o valor observado do genótipo mais correlacionado no ambiente em que não foi

possível a avaliação do genótipo não-completo. O valor resultante da variável dependente é a média estimada desejada.

Assim, o modelo estatístico adotado é:

$$Y_{ij} = \beta_0 + \beta_1 Z_{ij} + \delta_{ij} + \bar{\epsilon}_{ij}$$

em que  $Z_{ij}$  são os valores do genótipo  $i'$  de mais alto(a) coeficiente de determinação/correlação com o genótipo  $i$ .

### **3.1.6 Método VI – Manutenção da SQ (GxA)**

Essa metodologia se baseia na manutenção da soma de quadrados original da interação genótipos x ambientes após a estimação dos valores faltantes. Ou seja, por meio desse método, é estimada a média faltante que fornece para o experimento uma soma de quadrados de interação igual à soma de quadrados do experimento incompleto.

Sampaio (1998) relata um método de estimação de parcelas perdidas em experimentos sob o delineamento em blocos casualizados similar ao adotado neste trabalho. Esta estimação minimiza a soma de quadrados de erro do experimento. Entretanto, sabe-se que ao acrescentar uma parcela a uma soma de quadrados, ela nunca diminui; podendo ela, no mínimo, manter-se inalterada. Portanto, conclui-se que a metodologia citada em Sampaio (1998) também mantém a soma de quadrados de erros inalterada após a estimação da parcela perdida.

Como já exposto neste trabalho, a interação genótipos x ambientes se dá pela mudança de desempenho relativo de um genótipo frente às variações ambientais; ou seja, em um experimento conjunto sem qualquer interação, a diferença de rendimento médio de um genótipo em relação a outro é o mesmo em qualquer ambiente. Gráficamente, as curvas que representam os desempenhos de cada genótipo seriam paralelas e, portanto, nunca se cruzariam; o que, na prática, indica que a recomendação de um genótipo não dependeria do ambiente ou grupo de ambientes. Para ilustrar essa situação e facilitar o entendimento da filosofia desta metodologia de imputação, segue na Tabela 1 abaixo, o rendimento médio

de quatro genótipos em quatro ambientes sem qualquer interação genótipos x ambientes.

**Tabela1.** Exemplo ilustrativo de um experimento sem interação genótipos x ambiente.

Genótipos	Ambientes				Total
	A1	A2	A3	A4	
1	3	5	7	9	24
2	5	7	9	11	32
3	5	7	9	11	32
4	6	8	10	12	36
Total	19	27	35	43	124

Ao analisar as médias dos quatro genótipos nos quatro ambientes (Tabela1), é verificado que não há mudança de desempenho relativo dos genótipos frente às variações ambientais. Nota-se que a diferença de rendimento médio entre dois genótipos é a mesma em qualquer ambiente, isto denota nenhuma interação genótipos x ambientes.

Agora, para efeito de entendimento, ao tratar os ambientes acima como blocos e as médias como observações dos quatro genótipos nestes blocos, o experimento conjunto tornar-se-ia um ensaio de um fator sob o delineamento em blocos casualizados com quatro repetições, e este teria as seguintes somas de quadrados:

$$SQ(\text{Trat}) = \left( \frac{24^2 + 32^2 + 32^2 + 36^2}{4} \right) - \frac{124^2}{16} = 19$$

$$SQ(\text{Blocos}) = \left( \frac{19^2 + 27^2 + 35^2 + 43^2}{4} \right) - \frac{124^2}{16} = 80$$

$$SQ(\text{Total}) = (3^2 + 5^2 + 7^2 + 9^2 + 5^2 + 7^2 + 9^2 + 11^2 + 5^2 + 7^2 + 9^2 + 11^2 + 6^2 + 8^2 + 10^2 + 12^2) - \frac{124^2}{16} = 99$$

$$SQ(\text{Resíduo}) = 0$$

Este resultado para o valor da soma de quadrados do resíduo revela que a interação genótipos x ambientes em um experimento conjunto funciona como o resíduo em um experimento com um fator sob o delineamento em blocos casualizados. Dessa forma, seguindo o algoritmo descrito em Sampaio (1998), a imputação de parcela perdida em um experimento para um fator sob o delineamento em blocos casualizados manteria a soma de quadrado do resíduo. Logo, este algoritmo, usando a tabela de médias em experimentos conjuntos, também mantém a soma de quadrados da interação genótipos x ambientes.

O algoritmo de estimação descrito em Sampaio (1998) é o seguinte: substitui-se cada parcela (média) perdida por letras e procede-se à análise de variância literalmente. O valor da  $SQ(GxA)$  conterà todas as letras correspondentes às parcelas (médias) perdidas. Daí, deriva-se a  $SQ(GxA)$  em relação à cada incógnita separadamente e iguala-se cada equação obtida a zero. Feito isso, são encontradas  $p$  equações com  $p$  incógnitas, em que  $p$  indica o número de parcelas (médias) faltantes no ensaio. Finalmente, soluciona-se o sistema  $p \times p$  usando álgebra matricial, obtendo-se a resolução simultânea de todas as parcelas (médias) perdidas.

### **3.1.7 Ponderações sobre os métodos de imputação**

Nota-se que o método I é o mais simples dentre os seis elaborados. Com certeza, tal método de imputação dificilmente será utilizado na prática, entretanto, é importante que ele seja executado neste trabalho a fim de que se possa verificar a eficácia dos outros métodos propostos em relação a ele.

A segunda metodologia elaborada adiciona ao modelo estatístico adotado no método I o efeito de ambiente, porém ele pressupõe ausência de interação significativa em seu modelo; o que vai de encontro com a pressuposição da realização das análises de estabilidade e adaptabilidade.

Há, nas metodologias III, IV e V, dois problemas comuns que são dignos de comentários. Primeiramente, para a estimação das equações de regressão linear, é necessário que se estime os índices ambientais, que são as variáveis independentes para os métodos III e IV; e também são estimadas, por meio dos dados experimentais, as médias do genótipos do ensaio, estas são as variáveis independentes para o método V. Contudo, os valores das variáveis independentes

nestes métodos são estimativas, e, portanto, estão sujeitas a erros. Assim, Cruz e Regazzi (1997) afirmam que temos, nessa situação, um problema estatístico, que é o fato de o método de mínimos quadrados não prover um estimador consistente do coeficiente de regressão linear quando a variável independente estiver associada a erros.

O segundo problema comum a essas três metodologias é o fato de ao se realizar a estimação do valor faltante, pode-se haver uma extrapolação da equação de regressão estimada. A extrapolação pode ocorrer, pois tanto o valor de índice ambiental quanto o de média do genótipo utilizados na estimação desejada podem ser os maiores ou menores dentre os avaliados no ensaio. Portanto, a estimação de médias por meio desses valores não seria recomendável.

Ao se comparar os métodos III e IV, conclui-se que os índices ambientais do método IV são mais precisos que os estimados pelo método III, uma vez que eles são calculados por meio de médias dos mesmos genótipos. Porém, havendo um número grande de médias faltantes no experimento, o método IV torna-se impraticável, pois ele levará em consideração um número cada vez menor de médias observadas. Nestes casos, o uso da metodologia III seria mais adequado.

A metodologia V tem a vantagem de levar em consideração apenas o genótipo que mais se assemelha àquele no qual não houve avaliação em algum local. Entretanto, caso não haja também a avaliação do genótipo mais correlacionado a ele neste mesmo ambiente, não é possível proceder à estimação da média faltante. Uma alternativa menos precisa seria utilizar, nestes casos, o segundo genótipo mais correlacionado ou um conjunto de genótipos com um grau de correlação satisfatório.

A metodologia VI mantém a soma de quadrados da interação genótipos x ambientes. Tal manutenção significa que aquele valor não avaliado receberia um efeito de interação igual à média dos efeitos provenientes dos valores de médias avaliados no ensaio. Contudo, na prática, o que ocorre normalmente é que essa interação atua com efeito positivo ou negativo de maneira diferente nos diversos genótipos estudados.

### 3.2 Critérios de comparação das metodologias de imputação

Como já mencionado, além deste trabalho ter por finalidade desenvolver metodologias de imputação de médias faltantes em experimentos conjuntos incompletos, ele também tem o objetivo de avaliá-las.

Para que esse objetivo pudesse ser atingido, foi necessário que os valores a ser imputados fossem previamente conhecidos. Para tanto, foram obtidos dados experimentais relativos a um experimento conjunto completo de clones de café conilon, em dois locais, por nove colheitas, cedido pelo Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER). Esse conjunto de dados obtido é proveniente de um experimento conjunto com dados reais. É importante que se mencione o fenômeno da bienalidade, que ocorre com a cultura do café. Tal fato se caracteriza pela alternância anual entre alta e baixa produtividade da planta.

De posse desse conjunto de dados proveniente de um experimento real, foram retiradas aleatoriamente médias do ensaio, simulando, dessa forma, experimentos com 1%, 5% e 10% de médias faltantes. Para cada um destes três cenários, foram realizadas 10 réplicas. Ou seja, foram simulados trinta ensaios incompletos, dez com 1% de médias faltantes, dez com 5% de médias faltantes e os outros dez com 10% de médias faltantes.

Assim, cada um destes trinta ensaios incompletos tornaram-se completos por meio da imputação de médias faltantes mediante às seis metodologias elaboradas.

A fim de avaliar as metodologias de imputação de dados, foram estipulados três critérios, a saber: o erro quadrático médio, a correlação de Spearman entre a recomendação padrão do método de Lin & Binns (1988) e tal recomendação após a imputação das médias faltantes, e a porcentagem de mudanças na recomendação dos genótipos segundo à metodologia de Eberhart e Russel (1966).



### 3.2.1 Erro Quadrático Médio

Uma das maneiras de verificar qual dos métodos de estimação gera valores de médias que mais se aproximam dos valores reais em cada cenário é por meio do cálculo do erro quadrático médio para cada metodologia, que se baseia no quadrado da diferença entre o valor estimado e o valor observado. A metodologia que obtiver menor valor para esta estatística é o mais indicado para ser utilizado no cenário em questão por este critério.

O erro quadrático médio é calculado pela seguinte expressão:

$$EQM = \sum_{k=1}^r \frac{(Y'_{ijk} - Y_{ijk})^2}{r}$$

Em que:

$Y'_{ijk}$  = valor da média imputado do genótipo  $i$  no ambiente  $j$  na réplica  $k$ ;

$Y_{ijk}$  = valor observado da média do genótipo  $i$  no ambiente  $j$  na réplica  $k$ .

$r$  = número de réplicas.

Ao analisar a expressão que quantifica o erro quadrático médio, observa-se que este foi calculado para cada genótipo em cada um dos três cenários (1%, 5% e 10% de médias faltantes). Desse modo, foram comparados os erros quadráticos médios dos seis métodos de imputação em cada um desses cenários.

### 3.2.2 Coeficiente de correlação de Spearman

Outro critério de avaliação das metodologias de imputação é por meio do cálculo do coeficiente de correlação de Spearman (não-paramétrico).

A metodologia de estabilidade e adaptabilidade de Lin & Binns, como mencionado anteriormente, gera um valor da estatística  $P$  para cada genótipo. O menor valor para esta estatística indica que o genótipo é o mais recomendável para toda a rede experimental dentre os genótipos avaliados.

Ao submeter um conjunto de dados a esta metodologia de estabilidade e adaptabilidade por meio do *software* GENES, é fornecida a classificação dos

genótipos em ordem crescente para a estatística  $P$ . Assim, foram calculados os coeficientes de correlação de Spearman entre a classificação padrão dos genótipos, oriunda da análise de estabilidade e adaptabilidade do experimento conjunto completo, e a classificação após a imputação das médias faltantes.

A expressão que calcula o coeficiente de correlação de Spearman é a seguinte:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Em que:

$d_i$  = diferença entre a posição do genótipo  $i$  ao analisar as duas classificações (recomendação padrão e recomendação após a imputação das médias faltantes);

$n$  = número de genótipos avaliados no experimento.

Da mesma forma que o erro quadrático médio, os seis métodos de imputação foram comparados mediante o cálculo do coeficiente de correlação de Spearman em cada um dos três cenários estudados.

### 3.2.3 Modificações nas recomendações de Eberhart e Russel

Um terceiro critério utilizado para comparar as seis metodologias de imputação elaboradas foi a porcentagem de mudanças na recomendação dos genótipos advinda da metodologia de estabilidade e adaptabilidade de Eberhart e Russel.

A metodologia de Eberhart e Russel, como já explicado, calcula as estimativas dos parâmetros de estabilidade e adaptabilidade para cada genótipo. Em seguida, a metodologia testa as hipóteses de nulidade (1)  $H_0: \beta_{1i} = 1$  e (2)  $H_0: \sigma_{di}^2 = 0$ .

O parâmetro  $\beta_{1i}$  é estimado da seguinte maneira:

$$\hat{\beta}_{1i} = \frac{\sum_j Y_{ij} I_j}{\sum_j I_j^2}$$

Em que:

$Y_{ij}$  = média do genótipo  $i$  no ambiente  $j$ ;

$I_j$  = índice ambiental;

$j$  = número de ambientes em estudo.

Segundo Cruz e Regazzi (1997), uma avaliação da hipótese de nulidade (1) é feita por meio do teste  $t$ , cuja estatística é dada por:

$$t = \frac{\hat{\beta}_{1i} - 1}{\sqrt{\hat{V}(\hat{\beta}_{1i})}}$$

a qual está submetida a um nível de significância  $\alpha$  e ao número de graus de liberdade do resíduo, obtidos na análise de variância conjunta.

A variância do estimador  $\hat{\beta}_{1i}$  é dada pela seguinte expressão:

$$\hat{V}(\hat{\beta}_{1i}) = \frac{1}{\sum_j I_j^2} \hat{\sigma}_\varepsilon^2$$

Em que:

$\hat{\sigma}_\varepsilon^2 = \text{QMR}/r$ , e:

$\text{QMR} = \hat{\sigma}^2$  = quadrado médio do resíduo da análise de variância conjunta;

$r$  = número de repetições que deram origem às médias submetidas às análises de estabilidade e adaptabilidade.

A hipótese de nulidade (2) é avaliada pela estatística F, dada por:

$$F = \frac{QMD_i}{QMR}$$

Tal estatística está associada a um nível de significância  $\alpha$  e a  $a-2$  e  $m$  graus de liberdade, sendo  $m$  o número de graus de liberdade do resíduo obtido na análise conjunta e  $a$ , o número de ambientes avaliados.

Ainda conforme Cruz e Regazzi (1997), o parâmetro de estabilidade ( $\sigma_{di}^2$ ) é estimado pelo método da análise de variância a partir do quadrado médio do desvio da regressão de cada genótipo ( $QMD_i$ ) e do quadrado médio do resíduo, isto é:

$$\hat{\sigma}_{di}^2 = \frac{QMD_i - QMR}{r}$$

Em que:

$$QMD_i = \frac{r}{a-2} \left[ \sum_j Y_{ij}^2 - \frac{Y_i^2}{a} - \frac{(\sum_j Y_{ij} I_j)^2}{\sum_j I_j^2} \right]$$

Após a realização dos testes, o método de Eberhart e Russel rejeita ou não as hipóteses de nulidade (1)  $H_0: \beta_{1i} = 1$  e (2)  $H_0: \sigma_{di}^2 = 0$ , concluindo então, ao nível de significância estipulado (5%, neste estudo), se determinado genótipo é estável (previsível), possui adaptabilidade geral, possui adaptabilidade a ambientes favoráveis ou desfavoráveis etc.

Portanto, em cada cenário, foram calculadas as porcentagens de mudanças nas tomadas de decisões dos dois testes acima citados para cada metodologia de imputação. Obviamente, segundo este critério, a metodologia mais adequada é aquela que obtiver menor número percentual de mudanças nas tomadas de decisão em relação às duas hipóteses de nulidade supracitadas.

Este terceiro critério foi de fundamental importância para que os objetivos deste trabalho fossem alcançados. É com a análise destes números percentuais que

se pode verificar a influência da imputação de médias faltantes em experimentos incompletos na recomendação dos genótipos avaliados no estudo.

Deve ser ressaltado que uma ou mais médias faltantes em um experimento conjunto extenso, avaliado em muitos locais ou anos, é (são) muito comum (ns) de existir (em), o que só torna ainda mais evidente a importância deste estudo.

### **3.3 O Experimento conjunto completo**

A fim de que fosse possível avaliar os métodos de imputação de médias faltantes e verificar a influência da imputação das médias na recomendação dos genótipos oriunda das análises de estabilidade e adaptabilidade, foi necessário obter um conjunto de dados proveniente de um experimento conjunto completo.

O conjunto de dados foi fornecido pelo programa de melhoramento genético de café conilon do INCAPER (Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural).

Neste conjunto de dados, foram avaliados 40 genótipos (clones) de café conilon em 2 fazendas, a saber: Fazenda Experimental de Sooretama/Incaper e Fazenda Experimental de Marilândia/Incaper, ambas localizadas em macro-ambientes distintos do norte do Estado do Espírito Santo; a primeira, situada no município de Sooretama-ES; e a segunda, localizada no município de Marilândia-ES. As colheitas avaliadas foram realizadas nos anos de 1996, 1998, 1999, 2000, 2001, 2003, 2004, 2005 e 2006. Ou seja, os 40 genótipos (clones) de café conilon foram avaliados em 18 ambientes (2 fazendas x 9 colheitas).

O experimento foi conduzido sob o delineamento em blocos casualizados com 4 repetições, sendo que a parcela foi formada por duas plantas, uma vez que se trata de clone. O espaçamento utilizado foi de 3,0 x 1,5 metros, perfazendo uma população de 2222 plantas por hectare. A variável resposta avaliada foi produtividade (sacas beneficiadas de 60 kg por hectare). Vale ressaltar que os ambientes retratados nesse experimento conjunto representam 80% do café conilon plantado no estado do Espírito Santo.

Como mencionado acima, o experimento estudou 40 genótipos. Entretanto, houve valores faltantes de média envolvendo dois genótipos. Ou seja, dois genótipos não puderam ser avaliados em alguns ambientes; tratava-se, então, de um experimento conjunto incompleto. Porém, uma vez que para alcançar os

objetivos do trabalho, era preciso obter um experimento conjunto completo, os dois genótipos supracitados foram excluídos do estudo. Diante disso, o experimento conjunto passou a envolver 38 genótipos e tornou-se completo.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise dos dados

Para o experimento em questão, foi realizada a análise de variância, cujos resultados encontram-se na Tabela 2 e Tabela 3.

**Tabela2.** Análise de variância conjunta da produtividade de grãos (sacas beneficiadas/hectare) de clones de café no experimento conjunto com dados reais.

FV	GL	SQ	QM	F
BLOCOS/AMB	54	10594,76	196,20	
GENÓTIPOS	37	148271,20	4007,33	2,52
AMBIENTES	17	884689,28	52040,55	265,24
GEN X AMB	327	519069,45	1587,37	3,04
RESÍDUO	1008	526583,91	522,40	
TOTAL	2735	2089208,60		
MÉDIA	48,75			
CV(%)	46,88			

**Tabela3.** Testes de hipóteses realizados para o experimento com dados reais.

FV	TESTE	F	PROB.
GENÓTIPO	QMG/QMGA	2,52	< 0,0001
AMBIENTE	QMA/QMB	265,24	< 0,0001
TRAT X AMB	QMGA/QMR	3,04	< 0,0001

Pela análise da ANOVA, verifica-se que o modelo estatístico adotado foi o seguinte:

$$Y_{ijk} = \mu + B/A_{ij} + G_i + A_j + GA_{ij} + E_{ijk}$$

Nota-se também que o grau de liberdade da interação genótipo x ambiente não é igual ao produto dos graus de liberdade desses dois fatores analisados isoladamente. Tal fato ocorreu, pois foi calculado um grau de liberdade do resíduo ajustado, visto que não houve homogeneidade de variâncias ao realizar a ANOVA sem o ajuste. Com o ajuste, o grau de liberdade do resíduo passou a ser igual a

1008, diferente do valor original igual a 1998. Este grau de liberdade gerou um quadrado médio do resíduo igual a 522,40 e um coeficiente de variação igual a 46,88%; diferentes dos valores do quadrado médio do resíduo e coeficiente de variação originado pela análise de variância sem o ajuste do grau de liberdade, que foram iguais a, respectivamente, 263,56 e 33,30%.

Ainda, ao analisar os resultados dos testes F para genótipos, ambientes e genótipos x ambientes; verifica-se que as três hipóteses de nulidade foram rejeitadas. Segundo Vencovsky e Barriga (1992), quando um conjunto de materiais genéticos for avaliado em um certo número de ambientes, e as conclusões das pesquisas se referirem apenas a estes materiais, o modelo adotado deverá ser fixo. Como os pré-requisitos dos autores citados são atendidos, neste experimento, adotou-se o efeito de genótipo como fixo e o efeito de ambiente como aleatório (modelo misto); e, desse modo, foram rejeitadas as seguintes hipóteses de nulidade:

$$(1) \quad H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_{36} = \mu_{37} = \mu_{38}.$$

$$(2) \quad H_0: \sigma_A^2 = 0.$$

$$(3) \quad H_0: \sigma_{GA}^2 = 0.$$

A rejeição da hipótese de nulidade (1) indica que há, pelo menos, uma média diferente das demais. Em termos práticos, no experimento conjunto analisado, pelo menos um genótipo (clone) de café conilon possui média diferente dos demais.

A rejeição da hipótese de nulidade (2) revela que há variabilidade na população de ambientes. Como tal efeito foi considerado aleatório, pode-se dizer que os 18 ambientes avaliados é uma amostra representativa de sua população; e, como há variabilidade dentre os ambientes estudados, também há variabilidade na população.

A rejeição da hipótese de nulidade (3) evidencia que genótipos e ambientes atuam de modo independente sobre a variável resposta produtividade (sacas beneficiadas/hectare). Isto mostra que a seleção de um clone de café é dependente do ambiente. Ou melhor, a seleção dos clones de café deve ser realizada para um ambiente ou grupo de ambientes. Tal inferência mostra a importância da realização dos estudos de estabilidade e adaptabilidade.



Uma vez que existem mais de uma dezena de métodos de estabilidade e adaptabilidade conhecidos na literatura, foi necessário que fossem escolhidas as metodologias mais adequadas para os fins deste trabalho.

A metodologia de Eberhart e Russel foi escolhida por ser muito conhecida, devido a sua expressiva utilização em pesquisas acadêmicas desde seu surgimento, em 1966. Segundo Pelúzio *et al.* (2008), dentre as metodologias de estabilidade e adaptabilidade empregadas, o método de Eberhart e Russel (1966) é o mais usado. A metodologia não-paramétrica de Lin & Binns, por sua vez, foi escolhida devido a sua simplicidade de uso e interpretação dos resultados. Além disso, tal metodologia fornece uma classificação dos genótipos em relação aos demais, permitindo assim, que fossem comparadas as estruturas globais da classificação dos genótipos antes e após a imputação das médias. Deve-se ressaltar que com o uso da metodologia de Eberhart e Russel, isto não pode ser feito, uma vez que as conclusões tomadas a partir dos testes realizados são para cada genótipo, sendo errôneo compará-los de acordo com os p-valores correspondentes.

Ao submeter os dados deste ensaio a essas duas metodologias, foram geradas as recomendações padrão de cada genótipo. As análises de estabilidade e adaptabilidade foram realizadas pelo programa GENES (CRUZ, 2006).

A recomendação padrão do experimento conjunto completo com dados reais segundo a metodologia de Eberhart e Russel realizada pelo programa GENES (CRUZ, 2006) está evidenciada nas Tabelas 4 e 5.

**Tabela4.** Estimativas de adaptabilidade dos 38 clones de café avaliados no experimento conjunto com dados reais, Incaper 2012.

Genótipo	Média ( $\hat{\beta}_0$ )	$\hat{\beta}_1$	$t(\beta_1 = 1)$	Prob.(%)
1	41,5278	0,9673	-0,2184	82,2041
2	51,3333	1,1273	0,8497	59,9758
3	52,5278	1,1808	1,2068	22,5546
4	56,2333	0,8070	-1,2886	19,4640
5	48,1778	0,6192	-2,5423	1,0817
6	48,8833	1,0417	0,2787	77,7385
7	52,1222	0,8544	-0,9721	66,7213
8	51,3278	1,3699	2,4694	1,3173
9	55,0333	1,2019	1,3476	17,4415
10	43,0556	1,0526	0,3513	72,5709
11	46,4944	0,6715	-2,1932	2,6964
12	43,7278	0,8269	-1,1556	24,6597
13	29,4222	0,6719	-2,1904	2,7152
14	50,2278	0,9841	-0,1059	91,2210
15	47,7667	1,0849	0,5669	57,8185
16	54,9500	1,1448	0,9666	66,4358
17	48,8611	1,1157	0,7722	55,3743
18	35,4167	0,6585	-2,2797	2,1657
19	41,5667	0,8229	-1,1820	23,5568
20	47,9278	1,0480	0,3206	74,7316
21	46,1944	1,0736	0,4913	62,9161
22	53,6000	0,8293	-1,1394	25,3548
23	65,5000	1,7204	4,8088	0,0006
24	64,3000	1,2557	1,7072	84,0830
25	48,9556	0,7740	-1,5083	12,7460
26	39,5611	0,9789	-0,1406	88,3344
27	51,5000	1,1653	1,1033	26,9550
28	50,6167	1,0474	0,3162	75,0440
29	61,0667	1,0005	0,0032	99,2936
30	56,6000	1,0731	0,4879	63,1455
31	44,2556	0,8305	-1,1 311	25,7153
32	51,2778	1,0468	0,3125	75,3114
33	39,8222	0,5127	-3,2531	0,1353
34	38,5667	0,8607	-0,9297	64,4595
35	56,5056	1,1636	1,0918	27,4744
36	48,2056	1,1291	0,8619	60,6765
37	44,2889	1,2356	1,5728	11,1742
38	45,1444	1,0515	0,3439	73,0854

**Tabela 5.** Estimativas de estabilidade dos 38 clones de café avaliados no experimento conjunto com dados reais.

Genótipo	Média	S <sup>2</sup> d (1)	Prob.(%)	R <sup>2</sup> (%)
1	41,5278	0,9352	44,6153 NS	72,1235
2	51,3333	133,4960	0,9800 **	63,6387
3	52,5278	304,6904	0,0000 **	53,8110
4	56,2333	299,2787	0,0011 **	35,5243
5	48,1778	146,3525	0,6143 **	33,4863
6	48,8833	-72,6872	100,0 NS	87,2054
7	52,1222	29,5692	24,0283 NS	62,3728
8	51,3278	-10,1442	100,0 NS	85,0002
9	55,0333	45,4787	16,0369 NS	74,8987
10	43,0556	-45,7700	100,0 NS	82,6111
11	46,4944	-23,0442	100,0 NS	60,3905
12	43,7278	-90,6932	100,0 NS	86,1720
13	29,4222	-38,7901	100,0 NS	64,1365
14	50,2278	75,4542	6,7983 NS	63,0944
15	47,7667	66,4252	8,9101 NS	68,4832
16	54,9500	18,0776	31,3628 NS	76,2255
17	48,8611	69,7378	8,0774 NS	69,3240
18	35,4167	132,1154	1,0297 *	37,5129
19	41,5667	-45,0301	100,0 NS	74,2170
20	47,9278	-84,4098	100,0 NS	89,6363
21	46,1944	91,7346	4,0787 *	65,3451
22	53,6000	108,2006	2,3674 *	51,1607
23	65,5000	50,1032	14,1577 NS	85,6267
24	64,3000	232,4610	0,0208 **	61,2367
25	48,9556	-35,4018	100,0 NS	69,5970
26	39,5611	-57,9716	100,0 NS	82,7562
27	51,5000	29,9830	23,7889 NS	75,4636
28	50,6167	95,3583	3,6266 *	63,8442
29	61,0667	164,5519	0,3111 **	55,2271
30	56,6000	296,0996	0,0013 **	49,5352
31	44,2556	8,3777	38,5526 NS	64,3535
32	51,2778	62,3238	10,0422 NS	67,3837
33	39,8222	534,7485	0,0 **	12,5627
34	38,5667	-41,0385	100,0 NS	75,0541
35	56,5056	-40,3918	100,0 NS	84,5174
36	48,2056	54,9855	12,3720 NS	71,4175
37	44,2889	57,4385	11,5476 NS	74,7039
38	45,1444	-89,4871	100,0 NS	90,7254

(1) Valores negativos para as estimativas do componente de variância do desvio da regressão foram considerados nulos.

Nota-se, pela análise da Tabela 4, que o *software* fornece os valores estimados da constante de regressão  $\beta_0$  (média do genótipo) e do coeficiente de regressão  $\beta_1$ . Ademais, é realizado o teste *t* com a seguinte hipótese de nulidade:  $H_0: \beta_1 = 1$ . Em seguida, é fornecido o valor de probabilidade referente ao teste *t* executado. Este valor de probabilidade será comparado com o nível de significância adotado ( $\alpha = 5\%$ ) para que a tomada de decisão do teste em relação a rejeitar ou não rejeitar a hipótese de nulidade seja realizada.

Ainda na análise das estimativas de adaptabilidade, pode-se verificar que, para os genótipos um, dois, três, quatro e outros; as hipóteses de nulidade não foram rejeitadas, ou seja, os coeficientes de regressão  $\beta_1$  para esses clones foram considerados iguais a 1. Em contrapartida, para o genótipo cinco, oito e outros; a hipótese nula foi rejeitada; portanto, seus coeficientes de regressão foram considerados diferentes de 1.

Ao analisar a Tabela 5 referente às estimativas de estabilidade, nota-se que são fornecidas as médias dos genótipos, as estimativas das componentes de variância atribuídas aos desvios de regressão de cada genótipo, o valor de probabilidade associado à hipótese de nulidade  $H_0: \sigma_{di}^2 = 0$  e o valor do coeficiente de determinação  $R^2$  de cada regressão.

Segundo Cruz e Regazzi (1997), alguns genótipos com rendimento médio superior apresentam  $\sigma_{di}^2$  estatisticamente diferente de zero. Porém, pode ser necessária a seleção de alguns genótipos do grupo em que a estabilidade (previsibilidade) for baixa. Para tanto, uma medida auxiliar de comparação entre esses genótipos é o coeficiente de determinação  $R^2$ . É notado ainda que o software nos fornece a significância de cada estimativa de estabilidade. Desse modo, é percebido que o método considera o genótipo 1 como estável ou previsível, ao passo que identifica o genótipo 18 como pouco estável ou previsível ao nível de significância igual a 5%. Pode-se, ainda, inferir que um exemplo de genótipo recomendado a toda rede experimental seria o 16, uma vez que, segundo a metodologia utilizada, possui adaptabilidade geral ou ampla, alta estabilidade ou previsibilidade e possui alto rendimento médio (a média geral do experimento foi igual a 48,75).

Como já citado, a metodologia de estabilidade e adaptabilidade de Lin & Binns também forneceu uma recomendação padrão para este experimento

conjunto que também foi utilizada com a finalidade de avaliar as metodologias de imputação de médias faltantes elaboradas.

**Tabela6.** Estimativas de  $P_i$  da metodologia de Lin&Binns dos 38 genótipos de café avaliados no experimento conjunto com dados reais (sacas/hectare).

Genótipo	Média	$P_i$ Geral
23	65,5000	311,1711
24	64,3000	463,7528
29	61,0667	574,2106
35	56,5056	591,0264
9	55,0333	653,7600
16	54,9500	655,7275
8	51,3278	713,0914
30	56,6000	727,3733
4	56,2333	750,4767
22	53,6000	783,9022
7	52,1222	788,2744
27	51,5000	792,9789
28	50,6167	837,0875
6	48,8833	850,6597
32	51,2778	855,9394
2	51,3333	880,8278
3	52,5278	888,1369
36	48,2056	893,5192
14	50,2278	900,8675
20	47,9278	904,2003
17	48,8611	918,7581
25	48,9556	970,3083
15	47,7667	982,8356
38	45,1444	1021,2161
37	44,2889	1042,4394
21	46,1944	1080,8225
11	46,4944	1088,1736
10	43,0556	1120,7728
31	44,2556	1125,1794
12	43,7278	1131,0614
5	48,1778	1159,1700
1	41,5278	1205,6825
19	41,5667	1227,2239
26	39,5611	1297,8719
34	38,5667	1373,5033
18	35,4167	1703,1153
33	39,8222	1714,8939
13	29,4222	1893,1706

Ao analisar a Tabela 6, verifica-se uma significativa correspondência entre médias relativamente altas e valores relativamente pequenos para a estatística  $P_i$ . O genótipo 23 apresentou o menor valor para a estatística  $P_i$ , o que indica, pela metodologia de Lin & Binns, uma maior estabilidade e adaptabilidade geral aos ambientes avaliados no experimento. Coincidentemente, tal genótipo também possui a maior média dentre os genótipos avaliados no ensaio.

Deve-se frisar, porém, que nem sempre o genótipo com maior média geral apresentará menor valor para a estatística  $P_i$ . Esse fato já se evidencia ao analisar as produtividades (sacas beneficiadas/hectare) dos genótipos 8 e 30. O genótipo 8, apesar de apresentar menor média (51,3278) que o genótipo 30 (56,6000), possui um valor para a estatística  $P_i$  (713,0914) menor que do genótipo 30 (727,3733).

Como relatam Cruz e Carneiro (2003), vale ser lembrado que o conceito de estabilidade na metodologia de Lin & Binns retoma a ideia de invariância. Portanto, a estatística  $P_i$  torna-se eficaz na estimação dos parâmetros adaptabilidade e estabilidade fenotípica.

## **4.2 Avaliação dos critérios de comparação adotados**

### **4.2.1 Avaliação do Erro Quadrático Médio**

Como mencionado anteriormente, os seis métodos de imputação de dados elaborados foram avaliados segundo três critérios. O primeiro deles foi o Erro Quadrático Médio. Os valores correspondentes aos Erros Quadráticos Médios das seis metodologias de imputação nos três cenários estudados estão descritos na Tabela 7.

**Tabela 7.** Erros quadráticos médios das seis metodologias de imputação de dados nos três cenários avaliados.

Método	EQM(1%)	Método	EQM(5%)	Método	EQM(10%)
4	255,9 a	6	268,5 a	6	196,6 a
2	256,1 a	2	272,6 a	2	197,2 a
3	256,3 a	3	277,1 a	3	207,5 a
6	257,9 a	4	305,7 a	4	262,8 a
5	349,4 a	5	548,8 b	1	476,1 b
1	535,4 b	1	630,4 b	5	496,3 b

Métodos: (1)  $\mu + G_i$ ; (2)  $\mu + G_i + A_j$ ; (3) Regressão utilizando todos os dados disponíveis; (4) Regressão utilizando os dados dos genótipos completos; (5) Regressão com genótipo mais correlacionado; (6) Manutenção da soma de quadrados da interação.

Médias seguidas pela mesma letra não diferem entre si estatisticamente pelo teste de Tukey ao nível de 5% de probabilidade.

Os valores descritos na Tabela 7 devem ser interpretados da seguinte maneira: tal valor corresponde à média da diferença ao quadrado entre os valores imputados pelo método em questão e os respectivos valores reais em determinado cenário. Ou seja, o primeiro valor da Tabela (255,9) corresponde à média da diferença ao quadrado dos valores imputados pelo método 4 e os respectivos valores reais nos experimentos simulados com 1% de médias faltantes.

Visto que, no experimento estudado, havia 38 genótipos e 18 ambientes; foram geradas 684 médias. Assim, cada um dos 10 experimentos simulados com 1% de médias faltantes continham aproximadamente 7 perdas que foram determinadas aleatoriamente. O valor 255,9 é uma média de 68 valores de diferença ao quadrado.

Ao analisar os números da Tabela 7, podem-se constatar alguns fatos importantes. Primeiramente, observa-se a já esperada inferioridade do método 1 em relação aos demais métodos no cenário com 1% de médias faltantes. Nos cenários com 5% e 10% de médias faltantes, o método 5 também se mostra estatisticamente inferior aos demais (2, 3, 4 e 6) segundo o teste de Tukey ao nível de 5% de probabilidade.

Portanto, em experimentos conjuntos incompletos com 5% e 10% de médias faltantes, os métodos 2, 3, 4 e 6 se mostraram os mais adequados a serem utilizados no que se refere a imputar médias mais próximas das observadas; por outro lado, os métodos 1 e 5 se mostraram os menos adequados a esse fim.

#### 4.2.2 Avaliação da correlação de Spearman

O segundo critério de avaliação dos métodos de imputação elaborados foi a correlação de Spearman entre o *rank* dos valores da estatística  $P_i$  fornecido pela metodologia não-paramétrica de estabilidade e adaptabilidade de Lin & Binns e este mesmo *rank* realizado após a imputação dos dados.

Este valor de correlação indica em que grau a metodologia de imputação de médias faltantes influencia no aspecto global do *rank* dos genótipos quanto a sua adaptabilidade geral e estabilidade fenotípica. Abaixo, a Tabela 8 fornece os valores de correlação de Spearman das seis metodologias nos três cenários avaliados.

**Tabela 8.** Coeficientes de correlação de Spearman das seis metodologias nos três cenários avaliados

Método	$\hat{\rho}$ (1%)	Método	$\hat{\rho}$ (5%)	Método	$\hat{\rho}$ (10%)
3	0,9984 a	6	0,9897 a	6	0,9843 a
6	0,9982 a	3	0,9881 a	2	0,9837 a
2	0,9981 a	2	0,9871 ab	3	0,9825 a
4	0,9978 ab	4	0,9870 ab	4	0,9689 ab
5	0,9968 ab	5	0,9688 bc	1	0,9514 bc
1	0,9945 b	1	0,9615 c	5	0,9420 c

Métodos: (1)  $\mu + G_i$ ; (2)  $\mu + G_i + A_j$ ; (3) Regressão utilizando todos os dados disponíveis ; (4) Regressão utilizando os dados dos genótipos completos; (5) Regressão com genótipo mais correlacionado; (6) Manutenção da soma de quadrados da interação.

Médias seguidas pela mesma letra não diferem entre si estatisticamente pelo teste de Tukey ao nível de 5% de probabilidade.

Para que haja um maior entendimento do que foi realizado para que fossem obtidos os valores da Tabela 8, segue abaixo a Tabela 9, que mostra a recomendação padrão oriunda da metodologia de Lin & Binns e a recomendação após a imputação de médias em um experimento com 1% de médias faltantes (réplica 7) mediante o método 3 realizadas pelo programa GENES (CRUZ, 2006).



**Tabela9.** Recomendação padrão de Lin & Binns e após uma imputação de médias para os 38 clones de café mediante o método 3 no cenário com 1% de médias faltantes.

Genótipo	Média	$P_i$ (dados reais)	Genótipo	Média	$P_i$ (após imputação)
23	65,5	311,1	23	65,5	311,1
24	64,3	463,7	24	64,3	463,7
29	61	574,2	29	61	574,2
35	56,5	591	35	56,5	591
16	54,9	655,7	9	55	653,7
9	54,5	670,5	16	54,9	655,7
8	51,3	713	8	51,3	713
30	56,6	727,3	30	56,6	727,3
4	56,2	750,4	4	56,2	750,4
22	53,6	783,9	22	53,6	783,9
7	52,1	788,2	7	52,1	788,2
27	51,5	792,9	27	51,5	792,9
28	50,6	837	28	50,6	837
6	48,8	850,6	6	48,8	850,6
32	51,2	855,9	32	51,2	855,9
2	51,3	880,8	2	51,3	880,8
3	52,8	881,1	3	52,5	888,1
36	48,2	893,5	36	48,2	893,5
14	50,2	900,8	14	50,2	900,8
20	47,9	904,2	20	47,9	904,2
17	48,8	918,7	17	48,8	918,7
25	48,9	970,3	25	48,9	970,3
15	47,7	982,8	15	47,7	982,8
38	45,9	987,3	38	45,1	1021
37	44,2	1042	37	44,2	1042
21	46,8	1049	21	46,1	1080
11	46,4	1088	11	46,4	1088
10	43	1120	10	43	1120
31	44,2	1125	31	44,2	1125
12	43,7	1131	12	43,7	1131
5	48,1	1159	5	48,1	1159
1	41,5	1205	1	41,5	1205
19	41,5	1227	19	41,5	1227
26	39,5	1297	26	39,5	1297
34	38,5	1373	34	38,5	1373
18	35,8	1678	18	35,4	1703
33	39,8	1714	33	39,8	1714
13	29,4	1893	13	29,4	1893

As três primeiras colunas da Tabela 9 referem-se à recomendação padrão de Lin & Binns. As demais se referem à recomendação de Lin & Binns após a imputação de médias em um experimento com 1% de médias faltantes (réplica 7) mediante o método 3.

Nota-se grande concordância em relação ao *rank* dos genótipos quanto aos seus valores de  $P_i$ . Nesse caso, ao calcular o valor do coeficiente de Spearman entre os dois *ranks*, encontra-se o valor 0,99978. Pode-se perceber que apenas os postos dos genótipos 9 e 16 são modificados após tal imputação de dados, gerando, portanto, um valor de correlação próximo de 1.

O valor médio do coeficiente de correlação de Spearman para o método 3 no cenário de 1% é 0,9984 (Tabela 8). Tal valor é a média de 10 valores de correlação calculados em cada uma das réplicas simuladas para este método neste cenário específico.

Ao se analisar a Tabela 8, nota-se que, também para este critério, os métodos 1 e 5 apresentaram desempenho inferior em relação às metodologias 2, 3, 4 e 6 nos cenários com 5% e 10% de médias faltantes. Além disso, no cenário com 1% de perdas, as médias dos métodos 2, 3, 4, 5 e 6 foram consideradas iguais estatisticamente segundo o teste de Tukey ao nível de 5% de probabilidade. Deve-se ressaltar que os resultados obtidos neste critério se assemelham aos encontrados perante o critério anterior. Esta semelhança indica que valores imputados mais próximos dos valores reais tendem a gerar um *rank* também mais próximo do real.

### 4.2.3 Avaliação das modificações nas recomendações de Eberhart-Russel

Este critério, além de avaliar as metodologias elaboradas de imputação de médias faltantes, também verifica a influência desta imputação na recomendação dos genótipos.

A Tabela 10 mostra o número percentual de mudanças na recomendação dos genótipos das seis metodologias elaboradas nos três cenários estudados para o parâmetro de adaptabilidade.

**Tabela10.** Porcentagens de mudanças no parâmetro de adaptabilidade (PMPA) após a imputação das médias mediante os seis métodos elaborados nos três cenários estudados.

Método	PMPA(1%)	Método	PMPA(5%)	Método	PMPA(10%)
3	0,26	3	2,11	3	1,84
4	0,26	6	2,37	6	2,63
5	0,53	2	2,89	2	2,89
1	0,79	4	2,89	4	4,74
2	0,79	5	4,47	5	4,74
6	0,79	1	7,89	1	12,63

Métodos: (1)  $\mu + G_i$ ; (2)  $\mu + G_i + A_j$ ; (3) Regressão utilizando todos os dados disponíveis ; (4) Regressão utilizando os dados dos genótipos completos; (5) Regressão com genótipo mais correlacionado; (6) Manutenção da soma de quadrados da interação.

Como no critério anterior, para facilitar o entendimento sobre a obtenção dos dados da Tabela 10, segue abaixo a Tabela 11 que mostra a recomendação padrão de Eberhart e Russel e a recomendação após a imputação das médias mediante o método 1 em um experimento com 5% de médias faltantes (réplica 2).

**Tabela11.** Recomendação padrão de Eberhart e Russel e após uma imputação de médias mediante o método 1 no cenário com 5% de médias faltantes.

Genótipo	$\hat{\beta}_1$ (padrão)	Prob.(%)	Genótipo	$\hat{\beta}_1$ (imput.)	Prob.(%)
1	0,9673	82,2041	1	0,9848	91,9851
2	1,1273	59,9758	2	1,1851	23,5673
3	1,1808	22,5546	3	1,2627	8,9645
4	0,807	19,464	4	0,8568	63,6245
5	0,6192	1,0817	5	0,6378	1,9894
6	1,0417	77,7385	6	0,978	88,3572
7	0,8544	66,7213	7	0,9057	55,4866
8	1,3699	1,3173	8	1,0372	80,7566
9	1,2019	17,4415	9	1,2167	16,2792
10	1,0526	72,5709	10	1,0757	63,4679
11	0,6715	2,6964	11	0,6708	3,3777
12	0,8269	24,6597	12	0,6248	1,6036
13	0,6719	2,7152	13	0,7008	5,3316
14	0,9841	91,221	14	1,0008	99,1575
15	1,0849	57,8185	15	1,13	58,8488
16	1,1448	66,4358	16	1,224	14,8897
17	1,1157	55,3743	17	1,0966	54,5032
18	0,6585	2,1657	18	0,7022	5,4459
19	0,8229	23,5568	19	0,8681	59,5287
20	1,048	74,7316	20	1,0423	78,373
21	1,0736	62,9161	21	1,125	56,925
22	0,8293	25,3548	22	0,5942	0,9462
23	1,7204	0,0006	23	1,7581	0,0005
24	1,2557	8,4083	24	1,0754	63,6071
25	0,774	12,746	25	0,8109	22,5369
26	0,9789	88,3344	26	1,0292	84,6882
27	1,1653	26,955	27	1,2023	19,3509
28	1,0474	75,044	28	0,9997	99,4005
29	1,0005	99,2936	29	1,0411	78,8997
30	1,0731	63,1455	30	1,0966	54,5222
31	0,8305	25,7153	31	0,8764	56,4173
32	1,0468	75,3114	32	1,1067	50,333
33	0,5127	0,1353	33	0,5409	0,3587
34	0,8607	64,4595	34	0,8689	59,2535
35	1,1636	27,4744	35	1,1741	26,5877
36	1,1291	60,6765	36	1,1777	25,5575
37	1,2356	11,1742	37	1,2304	13,7324
38	1,0515	73,0854	38	1,0915	56,6402

Observa-se, pela análise da Tabela 11, que as estimativas dos coeficientes de regressão de todos os genótipos são diferentes antes e após a imputação das médias faltantes. Isso ocorreu em função de que tal estimativa é função dos índices ambientais, e estes se modificam após a imputação das médias faltantes.

Como já mencionado anteriormente, o nível de significância utilizado para a tomada de decisão em relação a rejeitar ou não as hipóteses nulas para o teste  $t$  ( $H_0: \beta_1 = 1$ ) igual a 5%. Desse modo, ao analisar a Tabela 11, verifica-se que a imputação dos dados mudou a tomada de decisão em relação ao parâmetro de adaptabilidade para os genótipos 8, 12, 13, 18 e 22, totalizando 5 mudanças dentre as 38 possíveis; o que representa uma mudança na recomendação de 13,16% dos genótipos. Na Tabela 10, encontra-se o valor 7,89% que é o percentual de mudanças na recomendação dos genótipos após a imputação das médias faltantes mediante o método 1 nos experimentos com 5% de perdas. Este valor é a média de dez valores referentes aos dez experimentos incompletos com 5% de médias faltantes simulados.

Ao analisar os resultados da Tabela 10, verifica-se também que nos cenários com 5% e 10% de perdas, o método 1 apresentou o pior desempenho. Nestes dois cenários, também houve concordância com os métodos que obtiveram os melhores desempenhos, sendo eles os de número 2, 3 e 6. O método 3 obteve melhor desempenho nos três cenários estudados para este critério, fato que pode ser explicado pelo modo semelhante ao de Eberhart e Russel que ele obtém os índices ambientais e, conseqüentemente, as estimativas dos coeficientes de regressão para cada genótipo.

Nos experimentos incompletos com 1% de médias faltantes, nenhum dos seis métodos de imputação elaborados atingiu 1% de mudanças na recomendação dos genótipos quanto à adaptabilidade.

À medida que o número de médias faltantes aumentou, cresceu também, quase que predominantemente, a porcentagem de erro na recomendação dos genótipos. Com exceção do método 1, nenhum método em nenhum cenário atingiu 5% de mudanças na recomendação dos genótipos.

A Tabela 12 mostra o número percentual de mudanças na recomendação dos genótipos das seis metodologias elaboradas nos três cenários estudados para o parâmetro de estabilidade.

**Tabela12.** Porcentagem de mudanças no parâmetro de estabilidade (PMPE) após a imputação das médias mediante os seis métodos elaborados nos três cenários estudados.

Método	PMPE(1%)	Método	PMPE(5%)	Método	PMPE(10%)
2	0,79	2	4,74	6	6,32
3	0,79	3	4,74	3	6,84
4	0,79	4	4,74	4	6,84
5	0,79	6	4,74	2	7,11
6	0,79	5	5,79	5	11,05
1	1,58	1	6,58	1	11,58

Métodos: (1)  $\mu + G_i$ ; (2)  $\mu + G_i + A_j$ ; (3) Regressão utilizando todos os dados disponíveis ; (4) Regressão utilizando os dados dos genótipos completos; (5) Regressão com genótipo mais correlacionado; (6) Manutenção da soma de quadrados da interação.

O valor para o método 1 no cenário com 1% de médias faltantes igual a 1,58% indica que dentre os 380 (38 genótipos X 10 réplicas) genótipos avaliados quanto a mudança ou não na tomada de decisão quanto à previsibilidade ou não do genótipo em questão, 6 (1,58%) deles tiveram sua recomendação quanto a sua estabilidade (previsibilidade) modificada; ou seja, genótipos cuja recomendação padrão indicavam como estáveis foram considerados instáveis ou vice-versa a 5% de nível de significância.

Pela análise da Tabela 12, é notado que as metodologias de imputação de médias faltantes apresentaram mesmo desempenho no cenário com 1% de perda, com exceção do método 1, que apresentou pior desempenho que os demais.

No cenário com 5% de médias faltantes, os métodos 2, 3, 4 e 6 apresentaram também mesmo desempenho, não atingindo 5% de mudanças na recomendação dos genótipos. Neste cenário, mais uma vez, os piores desempenhos ficaram com os métodos 1 e 5.

No cenário com 10% de médias faltantes, os métodos 2, 3, 4 e 6 apresentaram desempenhos relativamente semelhantes, enquanto que os métodos 1 e 5 quase atingem 12% de mudanças na recomendação dos genótipos.

Verifica-se também que os métodos de imputação modificam com mais frequência o parâmetro de estabilidade dos genótipos à medida que os cenários apresentam um maior número de médias faltantes.

## 5 CONCLUSÕES

Neste trabalho, foram propostas as seguintes metodologias de imputação de médias faltantes para fins de estudo de análise de estabilidade e adaptabilidade:

1 - A média imputada do genótipo  $i$  no ambiente  $j$  ( $Y_{ij}$ ) satisfaz o seguinte modelo matemático:  $Y_{ij} = \mu + G_i$ .

2 - A média imputada satisfaz o seguinte modelo matemático:  $Y_{ij} = \mu + G_i + A_j$ .

3 - A imputação se dá por meio de uma regressão linear simples em que a variável independente são os índices ambientais. Tais índices são estimados utilizando todos os dados disponíveis no experimento conjunto.

4 - A imputação se dá por meio de uma regressão linear simples em que a variável independente são os índices ambientais. Tais índices são estimados utilizando apenas os valores de médias dos genótipos que foram avaliados em todos os ambientes.

5 - A imputação se dá por meio de uma regressão em que a variável independente são as médias do genótipo de maior correlação com aquele que se deseja estimar o valor de média faltante.

6 - O método mantém a soma de quadrados da interação genótipo  $\times$  ambiente.

A variável dependente nos métodos 3, 4 e 5 são os valores de média observados do genótipo para o qual se deseja estimar um valor de média faltante.

As metodologias 2, 3, 4 e 6 apresentaram médias estatisticamente iguais em relação aos seguintes critérios: Erro Quadrático Médio e Coeficiente de correlação de Spearman. Elas são, portanto, as mais adequadas para a realização da imputação de médias faltantes em experimentos conjuntos incompletos dentre as metodologias desenvolvidas neste trabalho, visto que suas médias foram as maiores encontradas. Tais métodos imputaram médias mais próximas das observadas e foram as metodologias que menos alteraram a classificação geral dos genótipos quanto a sua adaptabilidade geral segundo o método de estabilidade e adaptabilidade de Lin & Binns (1988).



Conforme os resultados obtidos neste estudo, a realização da imputação das médias mediante as metodologias desenvolvidas com melhor desempenho em experimentos conjuntos incompletos foi aconselhável, uma vez que a alteração nas recomendações dos genótipos avaliados foi pequena se comparado com o número de médias faltantes nos ensaios avaliados.

## 6 REFERÊNCIAS

ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal of Genetics and Plant Breeding**, v. 46, p. 269-278, 1992.

ARCINIEGAS-ALARCÓN, S.; GARCIA-PENA, M.; DIAS, C. T. S. Imputação de dados em experimentos com interação genótipo  $\times$  ambiente. **Interciencia**, Caracas: v.6, n.6, p. 444-449, 2011.

ARCINIEGAS-ALARCÓN, S. *et al.* An alternative methodology for imputing missing data in trials with genotype by environment interaction. **Biometrical Letters**. v.47, p. 1-14, 2010.

ARCINIEGAS-ALARCÓN, S.; DIAS, C. T. S. Imputação de dados em experimentos com interação genótipo por ambiente: uma aplicação a dados de algodão. **Revista Brasileira de Biometria**. São Paulo: v.27, n.1, p. 125-138, 2009

BECKER, H.C.; LEON, L. Stability analysis in plant breeding. **Plant Breeding**, v.101, n.1, p.1-23, 1988.

BERGAMO, G.C.; DIAS, C.T.S.; KRZANOWSKI, W.J. Distribution-free multiple imputation in an interaction matrix through singular value decomposition. **Scientia Agricola**, Piracicaba, v.65, n.4, p.422-427, 2008.

BESPALHOK FILHO, J. C.; OLIVEIRA, R. A.; GUERRA, E. P. **Melhoramento de Plantas**. UFPR – Universidade Federal do Paraná, 2008. Disponível em: <[www.bespa.agrarias.ufpr.br](http://www.bespa.agrarias.ufpr.br)>. Acesso em: 17 jan. 2012.

BRASIL, E. M.; CHAVES, L. J. Utilización de um modelo cuadrático para el estudio de La resposta de cultivares a la variación ambiental. In: CONGRESSO LATIOMERICANO DE GENÉTICA, 11., 1994, Monterrey. **Memorias...** Monterrey: Asociación Latinoamericana de Genética, p. 616, 1994.

CALINSKI, T. *et al.* EM and ALS algorithms applied to estimation of missing data in series of variety trials. **Biuletyn Oceny Odmian**, Poznan, v.24-25, p.7-31, 1992.

CHAVES, L. J.; VENCOSKY, R.; GERALDI, I. O. Modelo não-linear aplicado ao estudo da interação genótipo  $\times$  ambiente em milho. **Pesquisa Agropecuária Brasileira**, v. 24, n. 2, p. 259-268, 1989.

CORNELIUS, P. L.; CROSSA, J.; SEYEDSADR, M. S. Statistical test and estimator of multiplicative model for genotype-by-environment interaction. In: KANG, M.S. e GAUCH Jr, H. G. (Ed.). **Genotype-by-Environment Interaction**. New York: Boca raton. p. 199-234, 1996.

- CRUZ, C. D. **Programa Genes**: Biometria. Editora UFV. Viçosa (MG). 382p. 2006
- CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. volume 2. Viçosa: UFV, 2003. 623p.
- CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa: UFV, 1997. 390p.
- CRUZ, C. D.; TORRES, R. A de.; VENCOSKY, R. An alternative approach to the stability na analysis proposed by Silva e Barreto. **Revista Brasileira de Genética**, Ribeirão Preto, v. 12, p. 567-580, 1989.
- DENIS, J. B.; BARIL C.P. Sophisticated models with numerous missing values: the multiplicative interaction model as an example. **Biuletyn Oceny Odmian**, Poznan, v.24-25, p.33-45, 1992.
- EBERHART, S. A; RUSSEL, W. A. Stability parameters for comparing varieties. **Crop Science** , Madison, v.6, p.36-40, 1966.
- FALCONER, D. S., MACKAY, T. F. C. **Introduction to quantitative genetics**. 4. ed. Edinburgh: Longman Group. 1998. 464 p.
- FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant breeding programme. Australian. **Journal of Agricultural Research**, v. 14, n. 6, p. 742-754, 1963.
- FOX, P. N. *et al.* Yield and adaptation of hexaploid spring triticale. **Euphytica**, v. 47, n. 1, p. 57-64, 1990.
- FREEMAN, H. G. Analysis of interactions in incomplete two-ways tables. **Applied Statistics**.v.24, p. 46-55, 1975.
- FRANCESCHI, L. *et al.* Métodos para análise de adaptabilidade e estabilidade em cultivares de Trigo no Estado do Paraná. **Bragantia**, v.69, n.4, p.797-805, 2010.
- GAUCH H. G, ZOBEL R. W. Imputing missing yield trial data. **Theoretical and Applied Genetics**,v. 79, p. 753-761, 1990.
- GODFREY, A.J.R; WOOD, G.R.; GANESALINGAM, S.; NICHOLS, M.A.; QIAO, C.G. Two-stage clustering in genotype-by-environment analyses with missing data. **Journal of Agricultural Science**, Cambridge, v.139, p.67-77, 2002.
- HUEHN, M. Nonparametric measures of phenotypic stability. Part 1 and 2: Theory and Applications. **Euphytica**, Wageningen, v. 47, n. 3, p. 189-194; 195-201, 1990.
- KANG, M. S. Using genotype-by-environment interaction for crops cultivar development. **Advances in Agronomy**, San Diego, v.62, p. 199-252, 1998.

- LAVORANTI, O. J. **Estabilidade e adaptabilidade fenotípica através da reamostragem “bootstrap” no modelo ammy**. Tese (Doutorado) - Universidade de São Paulo. Orientador: Carlos Tadeu dos Santos Dias, Piracicaba, SP, 2003. 166f.
- LIN, C. S.; BINNS, M. R. A superiority measure of cultivar performance for cultivar x location data. **Canadian Journal of Plant Science**, v. 68, p. 193-198, 1988.
- MAGARI, R.; KANG, M. S. SAS-STABLE: Stability analysis of balanced and unbalanced data. **Agronomy Journal**, v. 89, p. 929-932, 1997.
- MANDEL, J. The analysis of two-way tables with missing values. **Applied Statistics**. v.42, p. 85-93, 1993.
- MARIOTTI, J. A. *et al.* Analisis de estabilidad y adaptabilidad de genótipos de caña de azúcar, I, Interacciones dentro de una localidad experimental. **Revista Agronomicadel Noroeste Argentino**, Tucuman, v.13, n. 1-4, p.105-127, 1976.
- MOLINA, L. M. R. **Um estudo sobre métodos estatísticos na avaliação da interação genótipo x ambientes em genótipos de arroz (Oryza sativa L.)**. Dissertação (Mestrado) – Universidade Estadual Paulista. Orientador: Adhemar Sanches, Jaboticabal, SP, 2007. 79f.
- PERKINS, J. M.; JINKS, J. L. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. **Heredity**, Edinburgh, v.23, p. 525-35, 1968.
- PELUZIO, J. M. *et al.* Adaptabilidade e estabilidade de cultivares de soja em quatro épocas de semeadura no sul do Estado do Tocantins. **Revista Ceres**, v.55, n.1, p. 34-40, 2008.
- PLAISTED, R. L.; PETERSON, L. C. A technique for evaluating the ability of selection to yield consistently in different locations or seasons. **Am. Potato J.**, Washington, v. 36, n. 6, p. 381-5, 1959.
- ROCHA, M. M. *et al.* Estabilidade e adaptabilidade produtiva em linhagens de soja de ciclo médio. **Revista Ceres**, v.56, n.6, p. 764-771, 2009.
- ROSSE, L. N.; VENCOSKY, R. Modelo de regressão não-linear aplicado ao estudo da estabilidade fenotípica de genótipos de feijão no Estado do Paraná. **Bragantia**, Campinas, v. 59, n. 1, p. 99-107, 2000.
- SAMPAIO, I. B. M. **Estatística aplicada à experimentação animal**. 1ª Ed. Belo Horizonte: Fundação de Ensino e Pesquisa em Medicina Veterinária e Zootecnia, 1998. 221 p.
- SCHMILDT, E. R.; CRUZ, C. D. Análise de Adaptabilidade e Estabilidade do milho pelos métodos de Eberhart e Russel e de Annicchiarico. **Revista Ceres**, v.52, n.299, p. 45-58, 2005.

SILVA, J. G. C. Análise de adaptabilidade por regressão segmentada com estimação da junção dos segmentos. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 7, p. 1013-1029, 1998.

SILVA, J. G. C.; BARRETO, J. N. Aplicação da regressão linear segmentada em estudos da interação genótipo por ambiente. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 1., 1985, Piracicaba. **Anais...** Campinas: Fundação Cargill, p. 49-50, 1985.

SHUKLA, G. K. Some statistical aspects of partitioning genotype-environment components of variability. **Heredity**, Edinburgh, v. 29, n. 2, p. 237-45, 1972.

STORCK, L.; VENCOVSKY, R. Stability analysis based on a bi-segmented discontinuous model with measurement errors in the variables. **Revista Brasileira de Genética**, Ribeirão Preto, v. 17, n. 1, p. 75-81, 1994.

TAI, G. C. C. Genotypic stability analysis and its application to potato trials. **Crop Science**, Madison, v. 11, n. 2, p. 184-90, 1971.

TOLER, J. E.; BURROWS, P. M. Genotypic performance over environmental arrays: a non-linear grouping protocol. **Journal of Applied Statistics**, Abingdon, v. 25, n. 1, p. 131-143, 1998.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 496p.

VERMA, M. M.; CHAHAL, G. S.; MURTY, B. R. Limitations of conventional regression analysis, a proposed modification. **Theoretical and Applied Genetics**, Berlin, v. 53, n.1, p. 89-91, 1978.

YATES, F.; COCHRAN, W. G. The analysis of groups of experiments. **The Journal of Agricultural Science**, v. 28, p. 556-580, 1938.

WRICKE, O. Ubereinemethode zur erfassung der ökologischen streubreite in feldversuchen. **Z. Pflanzenzucht**, Berlin, v. 47, n. 1, p. 92-6, 1962.