

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA E MODELAGEM
QUANTITATIVA

Elisandra Lúcia Moro Stochero

**A INFLUÊNCIA DA VARIABILIDADE DOS DADOS NA
QUALIDADE DE IMPUTAÇÃO DE DADOS FALTANTES**

Santa Maria, RS
2019

Elisandra Lúcia Moro Stochero

**A INFLUÊNCIA DA VARIABILIDADE DOS DADOS NA QUALIDADE DE
IMPUTAÇÃO DE DADOS FALTANTES**

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Especialista em Estatística e Modelagem Quantitativa**.

Orientadora: Prof.^a Dr.^a Luciane Flores Jacobi
Coorientador: Prof. Dr. Alessandro Dal'Col Lúcio

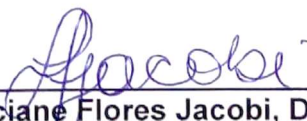
Santa Maria, RS
2019

Elisandra Lúcia Moro Stochero

**A INFLUÊNCIA DA VARIABILIDADE DOS DADOS NA QUALIDADE DE
IMPUTAÇÃO DE DADOS FALTANTES**

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Especialista em Estatística e Modelagem Quantitativa.**

Aprovado em 26 de março de 2019:



Luciane Flores Jacobi, Dra. (UFSM)
(Presidente/Orientadora)



Alberto Cargnelutti Filho, Dr. (UFSM)



Ana Lucia Souza Silva Mateus, Dra. (UFSM)

Santa Maria, RS
2019

RESUMO

A INFLUÊNCIA DA VARIABILIDADE DOS DADOS NA QUALIDADE DE IMPUTAÇÃO DE DADOS FALTANTES

AUTORA: Elisandra Lúcia Moro Stochero
ORIENTADORA: Prof.^a Dr.^a Luciane Flores Jacobi
COORIENTADOR: Prof. Dr. Alessandro Dal'Col Lúcio

Métodos de Imputação foram desenvolvidos com o objetivo de definir estimativas para dados ausentes em um banco de dados e, desta forma, solucionar possíveis problemas gerados pela perda de tais informações. Neste estudo o objetivo é avaliar se a variabilidade dos dados influencia nos resultados obtidos após aplicar um método de imputação. A partir de bancos de dados reais completos, de experimentos conduzidos no Delineamento em Blocos Casualizados, alguns com maior e outros com menor variabilidade, foram gerados bancos de dados incompletos com a retirada de diferentes quantidades de dados. Posteriormente, foi aplicado o método de Imputação Múltipla Livre de Distribuição, gerando bancos de dados completos a partir da imputação. Para os dados deste estudo, verificou-se que a variabilidade dos mesmos influenciou de maneira negativa quando alta e nos casos em que a variabilidade era baixa os valores imputados são mais próximos dos reais. Confirmando assim, a importância de avaliar a variabilidade dos dados antes de optar por aplicar o método de imputação.

Palavras-chave: Dados Ausentes. Imputação de Dados. Delineamento em Blocos Casualizados. Imputação Múltipla Livre de Distribuição.

ABSTRACT

THE INFLUENCE OF VARIABILITY OF DATA IN THE QUALITY OF MISSING DATA IMPUTATION

AUTHOR: Elisandra Lúcia Moro Stochero
ADVISOR: Prof.^a Dr.^a Luciane Flores Jacobi
CO-ADVISOR: Prof. Dr. Alessandro Dal'Col Lúcio

Imputation methods were developed with the purpose of defining estimates for missing data in a database and, in this way, solving possible problems generated by the loss of such information. In this study the objective is to evaluate if the variability of the data influences the results obtained after applying an imputation method. From complete real databases, from experiments conducted in the Randomized Block Design, some with larger and others with less variability, incomplete databases were generated with the withdrawal of different amounts of data. Subsequently, the Free Distribution Multiple Imputation method was applied, generating complete databases from the imputation. The results of the research confirm the importance of evaluating the variability of data before joining the application of an imputation method to obtain complete databases. For the data of this study, it was verified that the variability of the same influenced in a negative way when high and in cases in which the variability was low the imputed values are closer to the real ones. This confirms the importance of evaluating the variability of the data before choosing to apply the imputation method.

Keywords: Missing Data. Imputation of Data. Design in Randomized Blocks. Free Distribution Multiple Imputation.

LISTA DE FIGURAS

Figura 1 - Padrões de não resposta: (a) Padrão Univariado; (b) Padrão de Não Resposta Unitário; (c) Padrão Geral; (d) Padrão Monótono;	17
Figura 2 - Representação gráfica dos mecanismos de dados ausentes	19
Figura 3 – Representação gráfica da Imputação Múltipla	21

LISTA DE QUADROS

Quadro 1 – Variáveis avaliadas em cada banco de dados e notação adotada como referência para cada variável.....	29
Quadro 2 – Classificação do coeficiente de variação experimental.	31

LISTA DE TABELAS

Tabela 1 – Coeficiente de variação do conjunto de observações do banco de dados inicial e medidas de desempenho dos valores imputados.	32
Tabela 2 – Valores de precisão dos resultados após a retirada de diferentes valores de um mesmo banco de dados.	34
Tabela 3 – Coeficiente de variação experimental para os bancos completo, incompleto e completo com imputação para os dados do estudo.	34

SUMÁRIO

1. INTRODUÇÃO	10
1.1 OBJETIVOS	12
1.1.1 Objetivo Geral	12
1.1.2 Objetivos Específicos	12
1.2 JUSTIFICATIVA	12
1.3 ESTRUTURA DO TRABALHO	12
2. REVISÃO DE LITERATURA	14
2.1 DADOS FALTANTES	14
2.2 DESENVOLVIMENTO HISTÓRICO	15
2.3 PADRÃO E MECANISMOS DE NÃO RESPOSTA	16
2.4 MÉTODOS DE IMPUTAÇÃO DE DADOS FALTANTES	19
2.4.1 Método de Imputação Múltipla Livre de Distribuição (IMLD)	22
3. MATERIAL E MÉTODOS	29
4. RESULTADOS	32
5. CONCLUSÃO	36
REFERÊNCIAS	37
ANEXO A – BANCOS DE DADOS COMPLETOS, INCOMPLETOS E COMPLETOS COM VALORES IMPUTADOS	39
ANEXO B – VALORES DE PRECISÃO DOS RESULTADOS, RETIRANDO DIFERENTES VALORES DE UM MESMO BANCO DE DADOS	51
ANEXO C – ANÁLISE DAS 5 IMPUTAÇÕES PARA CADA DADO AUSENTE	57

1. INTRODUÇÃO

Vivemos em uma era em que diariamente são produzidas informações e um grande volume de dados. Junto a isso o processo de tomada de decisões com suporte dado pela Estatística também é crescente devido a confiabilidade e consistência dos resultados obtidos a partir de métodos estatísticos. Sua importância se ratifica pela sua aplicabilidade nas mais diversas áreas.

Porém, é comum deparar-se com alguns fatores que interferem na qualidade dos resultados de uma análise estatística, como a perda de informações durante o processo de coleta dos dados no decorrer do estudo. De acordo com Nunes, Klück e Fachel (2009), é comum a ocorrência de dados faltantes (*missing data*) em estudos científicos principalmente na área da Saúde e das Ciências Sociais.

Nunes, Klück e Fachel (2010) apontam como alguns exemplos que ocasionam a perda de informações o preenchimento incorreto de algum item, pacientes perdidos ao longo do estudo ou não-preenchimento de resultados de algum exame. Os autores também afirmam que problemas, como ineficiência e viés nas estimativas, podem ser gerados com a perda de dados.

Schafer e Graham (2002) levantam o questionamento sobre tais dificuldades criadas pelos dados faltantes na pesquisa científica, segundo os autores isso se deve ao fato de que a maioria dos procedimentos de análise de dados não foram assim projetados. A falta é geralmente um “contratempo” encontrado durante a pesquisa e não o foco principal, mas a maneira como tratado pode levantar dificuldades conceituais e desafios computacionais.

Como uma forma de contornar possíveis complicações geradas pela ausência de dados, de acordo com Dias e Albieri (1992), foram desenvolvidas técnicas de imputação de dados, onde são inseridas nas caselas vazias estimativas dos valores ausentes.

Para escolher entre o melhor método a ser aplicado, conforme Schafer e Graham (2002), é importante: ter conhecimento de alguns conceitos básicos; saber identificar o motivo pelo qual os dados estão em falta a fim de verificar a respeito do relacionamento entre a perda da informação e os valores existentes; analisar a forma como os dados faltantes estão distribuídos no banco de dados, ou seja, o padrão de dados ausentes.

Do ponto de vista prático, distinguir entre os padrões de dados ausentes não é tão importante, pois conforme Enders (2010), a estimativa da máxima verossimilhança e a imputação múltipla são adequadas para praticamente qualquer padrão de dados ausentes. Porém, durante o desenvolvimento histórico dos Métodos de Imputação, algumas técnicas analíticas desenvolvidas abordam um padrão específico de dados faltantes.

Os métodos de imputação podem ser divididos, de acordo com Engels e Diehr (2003), em: Imputação Única (IU), ou simples, que ocorre quando os dados ausentes são substituídos uma única vez e Imputação Múltipla (IM) que consiste em imputar vários valores para cada dado ausente obtendo-se para cada valor estimado um banco de dados completo que são avaliados aplicando-se diferentes métodos, após é determinada uma estimativa pontual de um parâmetro obtido através das imputações realizadas. O erro padrão é obtido através da variância das múltiplas imputações.

Embora o número de pesquisas neste contexto tenha sido crescente, principalmente na área da saúde, nas Ciências Agrárias onde os métodos estatísticos de análise e delineamento de experimentos são distintos não há muitos trabalhos considerando métodos de imputação.

Uma vez que a perda também pode ocorrer na experimentação agrícola, seja por rasuras, falta de preenchimento ou até mesmo perda de unidades experimentais, geralmente em pesquisas de campo, juntamente com a preocupação de obter resultados precisos e de qualidade, é interessante e de grande importância aprofundar e disseminar o estudo e aplicação de métodos de imputação.

Alarcón e Dias (2009) e Bergamo, Dias e Krzanowski (2008), por exemplo, trazem pesquisas mais recentes envolvendo as ciências agrárias no contexto de dados ausentes. Os dois estudos são direcionados para experimentos com interação Genótipo x Ambiente. Bergamo, Dias e Krzanowski (2008) inclusive propõem um novo método que abre mão das suposições referentes à distribuição dos dados ausentes.

No entanto, nenhum dos estudos tem como foco principal a influência que as características dos dados observados podem causar sobre os resultados dos métodos de imputação, sendo que para determinar as estimativas dos dados faltantes são utilizados os valores presentes no banco de dados.

Portanto, diante do exposto, o tema dessa pesquisa é verificar se a variabilidade dos dados provenientes de um experimento no Delineamento em Blocos Casualizados (DBC) influencia a qualidade da imputação dos dados perdidos.

1.1 Objetivos

1.1.1 Objetivo Geral

Avaliar a influência da variabilidade dos dados na qualidade de imputação de dados faltantes, comparando os resultados obtidos a partir da imputação entre bancos de dados com grande variabilidade e com menor variabilidade.

1.1.2 Objetivos Específicos

- Comparar os resultados entre bancos de dados com grande variabilidade e outros com menor variabilidade;
- Verificar se há influência da variabilidade dos dados na eficácia do método de imputação aplicado.

1.2 Justificativa

A análise estatística de dados tem sido aplicada cada vez mais e em diversas áreas como na saúde, nas ciências agrárias, na indústria, entre outras.

Um fato comum e que deve ser analisado cautelosamente é a ocorrência de dados ausentes. Esta ausência, de acordo com Schafer e Graham (2002), pode ter diversas naturezas que devem ser consideradas para se obter resultados corretos e não levar a inferências equivocadas.

Pesquisas referentes a dados faltantes e aplicações de métodos de imputação vem crescendo nas últimas décadas, conforme Nunes, Klück e Fachel (2009), e aprofundar esses estudos é de grande importância.

As consequências da análise incorreta dos dados, aplicações incorretas dos métodos para substituir dados ausentes por estimativas dos mesmos podem levar à resultados incorretos e estimativas com grande viés.

1.3 Estrutura do trabalho

Ao longo deste trabalho encontra-se um breve estudo referente aos Métodos de Imputação de Dados Ausentes já citados na Introdução.

No decorrer do Capítulo 2, é abordado os principais estudos referentes ao desenvolvimento histórico, conceitos, trabalhos desenvolvidos sobre imputação de dados e métodos. Sendo, neste trabalho, abordado detalhadamente um dos métodos de IM existentes.

Seguindo, no Capítulo 3 é descrita a metodologia adotada e os materiais disponíveis para dar continuidade ao estudo.

Resultados obtidos a partir da aplicação do método de imputação abordado no Capítulo 2 em bancos de dados reais são apresentados no Capítulo 4.

Também, é realizada a análise e comparação dos resultados descritos no Capítulo 4 com o objetivo de verificar se há possíveis interferências na qualidade e confiabilidade dos resultados após aplicar o método de imputação em relação a variabilidade dos dados iniciais. No Capítulo 5, é apresentada a conclusão deste estudo.

2. REVISÃO DE LITERATURA

2.1 Dados Faltantes

É comum, durante o processo de coleta de dados, ocorrer perda de informações, o que pode interferir nos resultados da análise dos dados ou até mesmo impossibilitar a aplicação de métodos estatísticos que exijam um banco de dados completo.

Conforme Schafer e Graham (2002), inicialmente é necessário avaliar o contexto, se realmente existe uma “informação verdadeira” por trás de cada casela vazia e se de fato é desconhecida. O autor destaca algumas possíveis situações que, se não considerada a forma como ocorreu a falta de determinado (s) dado (s), podem levar ao tratamento dos dados de maneira incorreta.

Por exemplo, em um levantamento de dados feito a partir de questionários existem algumas questões que estão interligadas, dependendo da resposta o participante segue respondendo sobre determinado assunto ou não, existe uma relação de dependência entre as variáveis. Porém, deve-se atentar ao fato de que o participante pode ter respondido incorretamente à questão que leva a resposta das outras. Neste caso, o pesquisador pode estar cometendo o erro ao considerar inexistente a resposta de todas as questões.

Uma outra situação envolvendo estudos longitudinais, o pesquisador apresenta o questionamento se, quando ocorrer eventos inesperados como a morte de um dos indivíduos, as informações que seriam avaliadas posteriormente a este evento seriam dados ausentes. O pesquisador, neste caso, pode simplesmente restringir a pesquisa considerando os indivíduos vivos, caso as mortes sejam raras e não relacionadas ao fenômeno de interesse.

Nas Ciências Agrárias, Banzatto e Kronka (2006) relatam sobre a possibilidade de chegar ao fim de um experimento e se deparar com a perda do valor de uma ou mais parcelas, as quais ele chama de parcelas perdidas. Entre as possíveis causas os autores citam: morte ou doença de plantas; falha do experimentador na coleta dos dados; extravio das anotações feitas sobre a parcela; valor muito discrepante dos demais que deve ser descartado; a parcela apresenta valor muito duvidoso.

Porém, é necessário ter cautela ao definir se realmente trata-se de um dado faltante. Em um estudo cujo objetivo é colher os frutos de cada parcela, por exemplo,

é possível deparar-se com três situações: a planta morreu; a planta não morreu, porém não há fruto; a planta não morreu, mas não há fruto apto a ser colhido. Tais situações não podem ser classificadas como dados faltantes, uma vez que a informação por trás da casela vazia não é desconhecida, ela é nula. Diferente da situação em que há a informação rasurada, onde existe uma informação verdadeira só que não é possível saber qual é devido a rasura.

A ocorrência de dados faltantes reduz o tamanho da amostra e pode nos levar a conclusões erradas. Tentativas de “corrigir” essa perda também pode afetar as inferências quando feita incorretamente. Portanto, para realizar novas pesquisas e entender melhor as já existentes é importante conhecer inicialmente como surgiram os Métodos de Imputação.

2.2 Desenvolvimento Histórico

Alguns procedimentos estatísticos foram desenvolvidos para trabalhar com conjuntos completos de dados o que gera dificuldade para o pesquisador quando encontradas caselas vazias. Inicialmente, a maneira mais comum era tratar estas situações considerando técnicas *ad hoc*, “manipulando” o conjunto de dados antes mesmo de realizar a análise.

Uma das técnicas *ad hoc* aplicadas é conhecida como exclusão *listwise* que, conforme Peugh e Enders (2004), é realizada a remoção geral dos casos com dados perdidos. Dependendo do tamanho da amostra e do número de variáveis, isso pode resultar em grande redução no tamanho da amostra disponível para análise de dados, (CARTER, 2006).

Técnicas de Imputação de Dados permitem preencher o banco de dados e assim ter uma matriz de dados completa, porém, há uma certa incerteza na validade dos resultados obtidos pelo fato de não serem os valores reais. Inicialmente, de acordo com Nunes, Kluck e Fachel (2009), os métodos aplicados eram simples, a substituição era feita pela média ou mediana da variável, por interpolação ou por regressão linear.

Os principais avanços em pesquisas envolvendo dados ausentes surgiram na década de 70 com a Estimção de Máxima Verossimilhança (EM) e Imputação Múltipla (IM), (ENDERS, 2010). Ainda, segundo o autor, Rubin foi responsável por estabelecer um sistema de classificação para problemas de dados ausentes.

Com o objetivo de encontrar as condições e dar suporte aos pesquisadores na decisão sobre considerar ou não o processo que causa a ausência de dados, Rubin (1976) apresenta exemplos destes processos, assim como, a distribuição condicional que corresponde ao processo gerador da ausência de dados em cada caso.

Little e Rubin (1987), apresentaram o método de Imputação Múltipla e desde então, segundo Schafer e Graham (2002), estudos relacionados a este tema vem crescendo, novos métodos foram desenvolvidos e a aplicação dos mesmos utilizadas em diversas áreas.

2.3 Padrão e Mecanismos de Não Resposta

Pelo fato de que muitas vezes estes termos são utilizados de maneira equivocada, se faz necessário saber quais são as diferenças entre mecanismo e padrão de dados ausentes.

O padrão de dados ausentes ou de não resposta, refere-se a maneira como estão dispostos os valores ausentes e observados em um conjunto de dados. De acordo com Enders (2010), o padrão de dados ausentes descreve a localização das caselas vazias não explicando o motivo desta falta.

Explanando os resultados em uma forma retangular ou matricial, onde linhas representam unidades observacionais ou participantes e colunas correspondem a itens ou variáveis, temos que, o padrão é dito **monotômico** quando ocorre dados ausentes em apenas uma das variáveis e **não-monotômico** quando ocorre dados ausentes em mais de uma das variáveis.

Na Figura 1, onde as áreas sombreadas representam a localização dos valores em falta, são apresentadas as classificações dos padrões mais citados na literatura, baseados nas representações expostas por Schafer e Graham (2002) e Enders (2010).

Padrão Univariado (*Univariate Pattern*): Os valores ausentes ocorrem em uma variável, mas um conjunto de outras variáveis é completamente observado. Em estudos experimentais é comum a ocorrência deste tipo de padrão.

Padrão de não resposta unitário (*Unit Nonresponse Pattern*): Encontrado em pesquisas que envolvem questionários, onde o indivíduo responde a somente algumas questões.

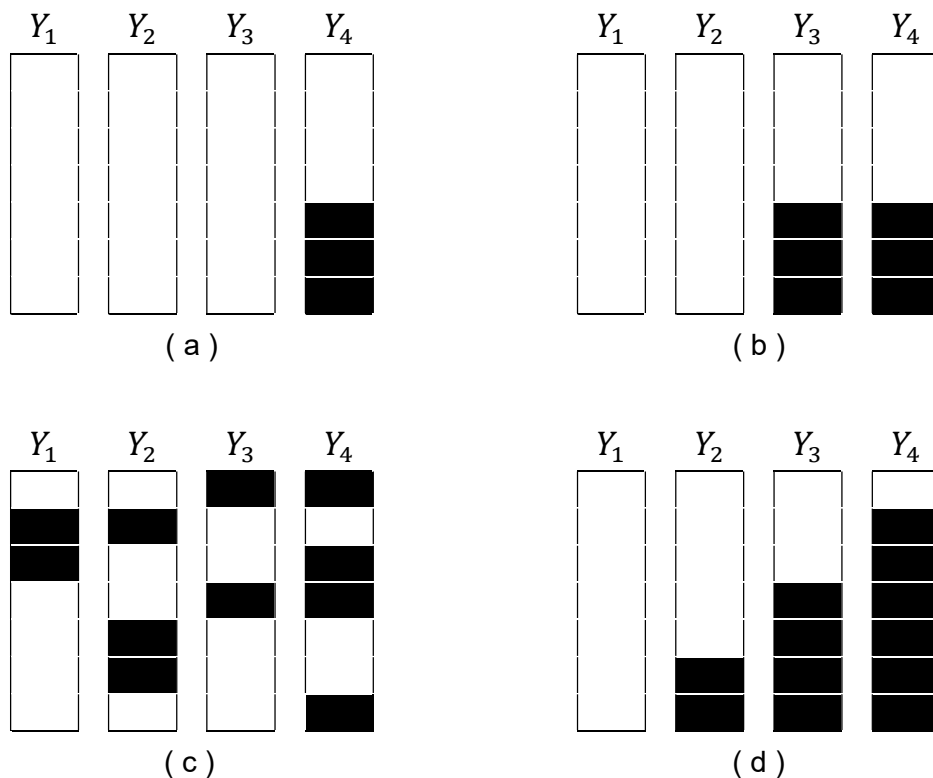
Padrão Geral (General Pattern): qualquer conjunto de variáveis pode apresentar dados em falta para qualquer unidade.

Padrão Monótono (Monotone Pattern): Itens ou grupos de itens Y_1, \dots, Y_p podem ser ordenados de tal forma que se Y_j estiver faltando para uma unidade, então Y_{j+1}, \dots, Y_p também estão ausentes. Padrões monótonos podem surgir em estudos longitudinais, quando algum participante não pode continuar no estudo.

De acordo com as definições, nas ciências agrárias, em um Delineamento Inteiramente Casualizado, a ordem da coleta das observações dentro de cada variável não interfere no resultado, sendo assim, possível reorganizar a maneira como estão dispostos os dados ausentes e observados deixando o banco de dados conforme o Padrão de Não Resposta Unitário.

Porém, no Delineamento em Blocos Casualizados, não é avaliado somente se há diferença significativa entre as variáveis, mas também entre os blocos. O padrão que melhor descreve a maneira como estariam dispostos os dados ausentes seria o Padrão Geral.

Figura 1 - Padrões de não resposta: (a) Padrão Univariado; (b) Padrão de Não Resposta Unitário; (c) Padrão Geral; (d) Padrão Monótono;



Fonte: Schafer e Graham (2002) e Enders (2010).

No entanto, quando se fala em mecanismos de dados ausentes, o objetivo é tentar inferir sobre o processo causador da falta de dados através da análise da distribuição dos dados ausentes, considerando a probabilidade de cada valor ser ou não observado (RUBIN, 1976).

Seguindo o sistema de classificação descrito por Rubin (1976), Schafer e Graham (2002) apresentam os tipos de mecanismos adotando, como o próprio autor diz, uma notação genérica.

Neste caso, de uma maneira mais simples, considera-se para qualquer conjunto de dados uma variável indicadora R que identifica o que é conhecido e o que está em falta. Onde, se o valor é observado, $R = 1$, caso o valor esteja ausente, $R = 0$. A variável R seria, assim, um conjunto de variáveis aleatórias com distribuição de probabilidade conjunta denominando-a como distribuição de falta ou probabilidade de ausência.

Denotando os dados completos como Y_{com} , e particionando como $Y_{com} = (Y_{obs}, Y_{miss})$ onde Y_{obs} refere-se a parte dos dados observados e Y_{miss} a parte dos dados ausentes, segue que os mecanismos de dados ausentes podem ser classificados da seguinte forma:

Dado Faltante Aleatório (*Missing at Random – MAR*): O padrão da perda pode ser predito através das outras variáveis do banco de dados e não especificamente à variável onde os dados estão faltando, ou seja, a probabilidade dos dados em falta depende dos dados observados, mas não dos dados ausentes;

$$P(R | Y_{com}) = P(R | Y_{obs}) \quad (1)$$

Dado Faltante Completamente Aleatório (*Missing Completely at Random – MCAR*): Este é um caso especial de MAR onde, as razões das perdas estão relacionadas a todos os dados presentes e inclusive os ausentes;

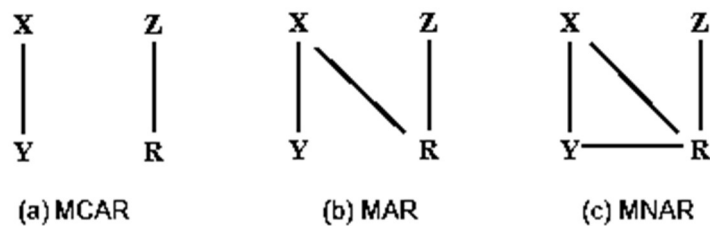
$$P(R | Y_{com}) = P(R) \quad (2)$$

Dado Faltante Não Aleatório (*Missing Not at Random – MNAR*): Este caso ocorre quando a condição do MAR é violada e a distribuição depende de Y_{miss} . Os dados não estão perdidos aleatoriamente.

Para facilitar a compreensão das classificações, Schafer e Graham (2002) desenvolveram uma representação gráfica, conforme pode ser visto na Figura 2. Onde, X = a variável em que todos os dados foram observados, Y = variável que apresenta dados ausentes e Z = componente das causas de ausência não relacionadas a X e a Y. R representa a falta completa, é a variável indicadora descrita anteriormente.

Na Figura 2(a), a ausência é completamente aleatória, a causa da ausência está ligada a outras variáveis diferentes de X e Y. Em 2(b) a ausência é aleatória, a causa não está relacionada a própria variável que possui dados ausentes e sim com fatores externos e a variável em que não há ausência de dados. Em 2(c), a ausência não é aleatória, e a causa da ausência está ligada a todas as variáveis, aquela com todos dados observados, a que possui dados ausentes e as que não foram avaliadas.

Figura 2 - Representação gráfica dos mecanismos de dados ausentes



Fonte: Schafer e Graham (2002)

2.4 Métodos de Imputação de Dados Faltantes

Ao deparar-se com um banco de dados incompleto, um dos procedimentos mais antigos adotados era a exclusão das unidades em que havia dados ausentes. Porém, este tipo de procedimento pode implicar em resultados com um grande viés, informações que influenciariam nos resultados podem acabar sendo descartadas além de que o tamanho da amostra também se torna menor.

A maioria dos trabalhos relacionados a imputação de dados são voltados para a área da saúde ou da psicologia, poucos são voltados a área das ciências agrárias, sendo menor ainda o número de publicações neste contexto direcionadas à experimentos desbalanceados.

Além de citar possíveis situações em que ocorre a perda de parcelas, como comentado anteriormente, Banzatto e Kronka (2006) indica uma maneira de lidar com

estas perdas, ou seja, o procedimento que pode ser adotado para calcular uma estimativa da parcela perdida. Com o banco de dados completo a análise é mais simples e não se faz necessária a utilização de métodos mais complexos na presença de blocos incompletos não balanceados.

Para a obtenção da estimativa do dado ausente para a parcela perdida x_{ij} correspondente ao tratamento i no bloco j , Banzatto e Kronka (2006) apresenta a seguinte expressão:

$$\hat{x}_{ij} = \frac{IT + JB - G'}{(I-1)(J-1)} \quad (3)$$

em que,

I = número de tratamentos do experimento;

T = soma das parcelas existentes no tratamento que teve a parcela perdida;

J = número de blocos do experimento;

B = soma das parcelas existentes no bloco que teve a parcela perdida;

G' = soma de todas as parcelas existentes no experimento.

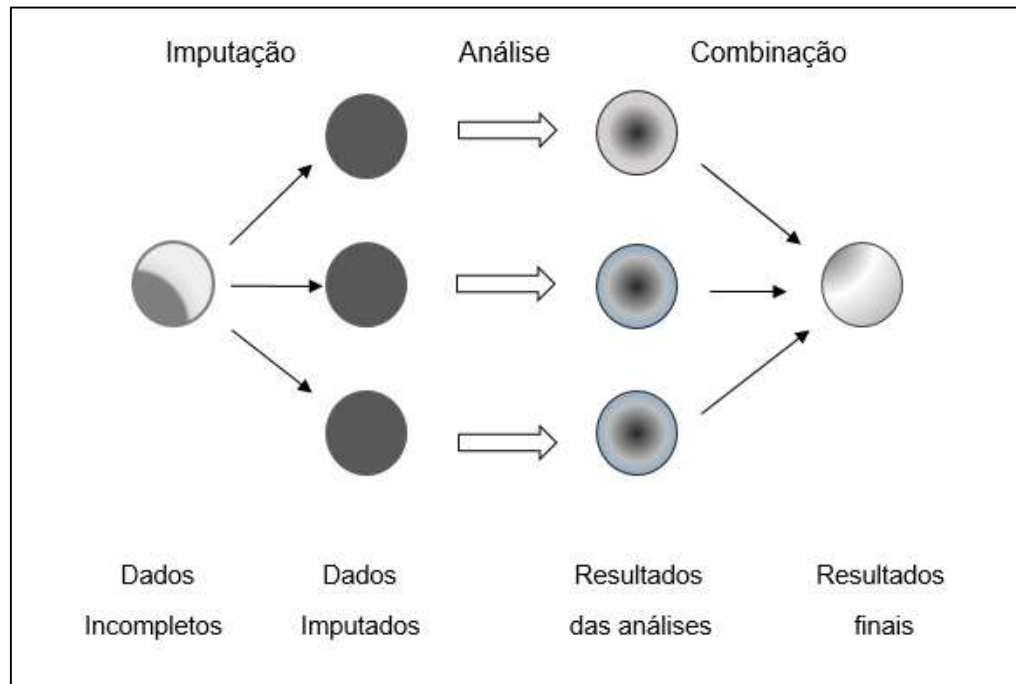
O método descrito acima é simples e na literatura é possível encontrar outros métodos, alguns também mais simples e outros mais complexos cuja aplicação é mais acessível atualmente devido ao avanço computacional.

A aplicação de métodos de imputação, como já mencionado, consiste em manter todos os dados observados e preencher as caselas vazias com estimativas dos valores perdidos, obtidas a partir dos valores conhecidos.

A imputação simples ou imputação única, como também é conhecida, consiste na inclusão do valor estimado no banco de dados uma única vez.

Na Imputação Múltipla, técnica proposta por Rubin para determinação da estimativa do valor a ser inserido no “espaço vazio” pode ser representado na forma gráfica como na Figura 3 e ser dividido nas seguintes etapas, conforme Nunes, Klück e Fachel (2009):

Figura 3 – Representação gráfica da Imputação Múltipla



Fonte: Nunes, Klück e Fachel (2010).

Etapa I: são criados m ($m > 1$) conjuntos de dados completos preenchendo os valores ausentes a partir de algum método de imputação;

Etapa II: os conjuntos com dados imputados são analisados como se fossem um conjunto de dados completo “real”;

Etapa III: os resultados da análise do conjunto de dados completos imputados são combinados de uma maneira simples e adequada.

O modelo utilizado na primeira etapa e na etapa em que é feita a análise podem ser distintos. Na etapa da análise são determinadas estimativas de parâmetro e erro padrão, dos m conjuntos de dados completos, posteriormente estas medidas são combinadas, conforme Enders (2006), através das equações (4), (5) e (6).

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i \quad (4)$$

em que, \hat{Q}_i é a estimativa dos parâmetros de cada conjunto de dados completos a partir da imputação.

Na combinação dos erros padrão são consideradas a variação dentro das imputações e entre as imputações a partir das equações (5) e (6), respectivamente.

$$\bar{U} = \frac{1}{m} \sum_{i=1}^m \hat{U}_i \quad (5)$$

$$B = \frac{1}{m-1} \sum_{i=1}^m (\hat{Q}_i - \bar{Q})^2 \quad (6)$$

em que, \hat{U}_i é a variância de cada conjunto de dados imputados.

Com as estimativas \bar{Q} , \bar{U} e B determinadas, o erro padrão combinado é dado pela equação (7).

$$SE = \sqrt{\bar{U} + \left(1 + \frac{1}{m}\right) B} \quad (7)$$

Neste trabalho é descrito o Método de Imputação Múltipla Livre de Distribuição proposto por Bergamo, Dias e Krzanowski (2008). O qual será utilizado para verificar se há diferença entre os resultados de bancos de dados com maior e menor variabilidade.

2.4.1 Método de Imputação Múltipla Livre de Distribuição (IMLD)

O método de imputação múltipla proposto por Bergamo, Dias e Krzanowski (2008) tem como ponto de partida a Decomposição por Valores Singulares (DVS) para imputação simples desenvolvido por Krzanowski em 1988. Os autores realizaram a avaliação do método proposto a partir de uma matriz de dados de um ensaio com delineamento aleatorizado em blocos, com genótipos de *Eucalyptus grandis*.

Sem fazer qualquer suposição referente a distribuição ou estrutura dos dados, foi aplicado a DVS da matriz “genótipo x ambiente” (G x A). A partir do experimento adotado para realizar a avaliação do método, foram feitas considerações iniciais sobre as notações adotadas.

Para descrever o método proposto, Bergamo, Dias e Krzanowski (2008) inicialmente mostram algumas considerações sobre a decomposição de uma matriz pela DVS, a qual serviu como base para determinar a dimensionalidade de um conjunto de dados multivariados.

A DVS consiste na fatoração de uma matriz qualquer $Y_{n \times p}$ da seguinte forma

$$Y = UDV^T \quad (8)$$

U: Matriz formada pelos autovetores de Y^TY ;

D: Matriz diagonal formada pelos autovalores de Y^TY .

V^T : Transposta da matriz formada pelos autovetores de YY^T .

em que, $U^TU = V^TV = VV^T = I_p$, $UU^T = I_n$ e $D = \text{diag}(d_1, \dots, d_p)$ com $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$. As matrizes Y^TY e YY^T possuem os mesmos autovalores e os elementos d_i são as raízes quadradas desses autovalores.

A i -ésima coluna $v_i = (v_{i1}, \dots, v_{ip})^T$ da matriz $V_{p \times p}$ é o autovetor que corresponde ao i -ésimo maior autovalor d_i^2 de Y^TY . Na matriz $U_{n \times p}$, a j -ésima coluna $u_j = (u_{1j}, \dots, u_{nj})^T$ é o autovetor que corresponde ao j -ésimo maior autovalor d_j^2 de YY^T . Logo, a decomposição da equação (8) pode ser dada da seguinte forma:

$$y_{ij} = \sum_{h=1}^p u_{ih} d_h v_{jh} \quad (9)$$

Krzanowski (1988 apud BERGAMO, DIAS E KRZANOWSKI, 2008), fizeram uso deste resultado para determinar a dimensão de um conjunto de dados multivariados concluindo que, se a estrutura de dados é essencialmente H-dimensional, então a variação nas dimensões restantes (p - H) pode ser tratada como ruído aleatório.

As principais características dos dados podem, portanto, estar no espaço dos primeiros H componentes principais. A correspondência entre as quantidades do lado direito da equação (9) e os eixos principais da configuração dos dados sugere, portanto, o modelo do componente H.

$$y_{ij} = \sum_{h=1}^H u_{ih} d_h v_{jh} + \varepsilon_{ij} \quad (10)$$

em que, ε_{ij} é um termo residual.

Supondo que o modelo (10) é válido para um valor específico de H, mas o valor referente a observação y_{ij} está ausente da matriz de dados, então y_{ij} pode ser estimado através da equação (11).

$$\hat{y}_{ij}^{(H)} = \sum_{h=1}^H u_{ih} d_h v_{jh} \quad (11)$$

Sendo que, u_{ih} , d_h e v_{jh} , são estimados a partir dos dados restantes.

As melhores estimativas dessas últimas quantidades serão aquelas que usam a quantidade máxima de dados. Denotando por $Y^{(-i)}$ a matriz de dados obtida ao deletar a i -ésima linha de Y , e por $Y_{(-j)}$ matriz de dados obtida ao deletar a j -ésima coluna de Y , e supondo que as DVS destas matrizes são dadas por:

$$Y^{(-i)} = \bar{U} \bar{D} \bar{V}^T, \bar{U} = (\bar{u}_{sh}), \bar{V} = (\bar{v}_{sh}), \bar{D} = (\bar{d}_1, \dots, \bar{d}_p) \quad (12)$$

e

$$Y_{(-j)} = \tilde{U} \tilde{D} \tilde{V}^T, \tilde{U} = (\tilde{u}_{sh}), \tilde{V} = (\tilde{v}_{sh}), \tilde{D} = (\tilde{d}_1, \dots, \tilde{d}_{p-1}) \quad (13)$$

As estimativas máximas de dados de u_{ih} e v_{jh} , encontradas a partir da equação (11), são \tilde{u}_{ih} e \tilde{v}_{jh} , respectivamente, enquanto d_h pode ser estimado por \bar{d}_h , \tilde{d}_h ou por alguma combinação dos dois. Segundo Krzanowski (1988 apud BERGAMO, DIAS E KRZANOWSKI, 2008), é aconselhado usar $\sqrt{\bar{d}_h} \sqrt{\tilde{d}_h}$ como um ajuste adequado, de modo que uma estimativa do valor omisso y_{ij} é dada pela equação (14).

$$\hat{y}_{ij}^{(H)} = \sum_{h=1}^H (\tilde{u}_{ih} \sqrt{\tilde{d}_h}) (\tilde{v}_{jh} \sqrt{\bar{d}_h}) \quad (14)$$

O valor mais alto possível de H deve ser usado, da equação (13) segue que este valor é $p - 1$, de modo que o valor a ser imputado por y_{ij} será:

$$\hat{y}_{ij} = \sum_{h=1}^{p-1} (\tilde{u}_{ih} \sqrt{\tilde{d}_h}) (\tilde{v}_{jh} \sqrt{\bar{d}_h}) \quad (15)$$

Para encontrar as quantidades na equação (15), é necessário um processo numérico iterativo. Começando com suposições iniciais para os valores ausentes, cada iteração requer DVS para as matrizes reduzidas $Y^{(-i)}$, $Y_{(-j)}$ para cada (i, j) onde há um valor ausente.

A utilização da equação (15) fornece a imputação atualizada para essa posição e o processo continua até ocorrer uma estabilidade nos valores sucessivos imputados. Inicialmente, são consideradas para os valores ausentes y_{ij} a média \bar{y}_j dos dados da j -ésima coluna em que se encontra a posição do dado faltante.

Para evitar qualquer variação possível na influência entre colunas, por exemplo, causada por diferentes escalas de medição, é recomendado primeiramente padronizar Y (matriz preenchida com as “suposições” iniciais).

As matrizes $Y^{(-i)}$ e $Y_{(-j)}$ também devem ser padronizadas em cada etapa, de modo que todas as operações sejam realizadas em quantidades padronizadas. No final, a matriz completa Y (valores observados + imputados) deve retornar à sua escala original. Assim, se $y^{(c)}_{ij}$ representa cada valor da matriz completa, a j -ésima média da coluna $\bar{y}^{(c)}_j$ e o desvio padrão $s^{(c)}_j$ são calculados e cada valor da matriz completa Y é obtido em sua escala original como $y_{ij} = \bar{y}^{(c)}_j + s^{(c)}_j y^{(c)}_{ij}$.

Para o método IMLD, Bergamo, Dias e Krzanowski (2008), seguem os passos anteriores e propõe uma generalização para os expoentes de \tilde{d}_h e \bar{d}_h presentes na equação (15), ao gerar as imputações ($m = 1, \dots, M$) no primeiro passo da IM.

Para a generalidade completa, se $\sqrt[b]{d^a}$ é representado como uma potência fracionária $d^{\frac{a}{b}}$, no método IMLD é realizada a mudança dos expoentes da equação (13) para $\tilde{d}_h^{\frac{a}{b}}$ e $\bar{d}_h^{\frac{a}{b}}$, de forma que a soma destes expoentes seja igual a 1 (isto é, $\frac{\tilde{a} + \bar{a}}{b} = 1$). Assim, diferentes pesos podem ser atribuídos aos resultados obtidos através das equações (12) e (13) na estimativa final de y_{ij} dada pela equação (15), variando os expoentes de \tilde{d}_h e \bar{d}_h , enquanto a forma atual os força a ter pesos iguais.

Cada valor diferente de \tilde{a} e conseqüentemente de \bar{a} gera uma nova matriz completa Y , fornecendo assim um mecanismo para gerar os M diferentes conjuntos de dados resultantes da primeira etapa do processo de IM.

O número de imputações tem relação direta com o número dos diferentes expoentes usados. Conforme, Schafer (1999 apud BERGAMO, DIAS E KRZANOWSKI, 2008) e Rubin (1987 apud BERGAMO, DIAS E KRZANOWSKI, 2008), entre 3 e 5 imputações deve ser suficiente para caracterizar a variabilidade entre as imputações. Assim, se for decidido em 5 mudanças nos expoentes, entre 40% e 60% a variação pode ser produzida no peso dado as equações (12) e (13) iniciando com um denominador fixo ($b = 20$, por exemplo) e tomando valores (8, 9, 10, 11 e 12) para \tilde{a} e (12, 11, 10, 9 e 8) respectivamente para \bar{a} . Estas escolhas levam a uma variação de (40%, 45%, 50%, 55% e 60%), respectivamente, nas proporções de (12) e (13).

Desta forma, a estimativa do valor ausente pode ser determinada através da equação (16).

$$\hat{y}_{ij} = \sum_{h=1}^{p-1} \left(\tilde{u}_{ih} \tilde{d}_h^{\frac{\tilde{a}}{b}} \right) \left(\bar{v}_{jh} \bar{d}_h^{\frac{\bar{a}}{b}} \right) \quad (16)$$

O procedimento proposto usa a maior quantidade de valores em Y e não depende de qualquer distribuição na variável de resposta, sendo aplicável a qualquer matriz de dados numéricos.

Para a aplicação do método, o autor fez uso de um programa desenvolvido no módulo IML do sistema estatístico SAS, que, após sua execução, resultou em um arquivo de dados com os $M = 5$ conjuntos de dados completos, prontos para serem utilizados na segunda etapa da IM.

Os procedimentos foram aplicados às três matrizes de dados reduzidos, contendo 5%, 10% e 30% de valores omissos. A estabilidade dos valores imputados, ou sua convergência para um valor único, foi geralmente alcançada em 20 iterações, mas, por razões de segurança e generalidade na execução do programa, 50 iterações foram a configuração padrão.

Como medida de desempenho do método na posição do valor omissos l (ausente na linha i e coluna j), foi utilizada a seguinte expressão:

$$acc_l = \frac{\sum_{m=1}^M (\hat{y}_{ij(m)} - VO_l)^2}{M-1} \quad (17)$$

VO_l é o valor original eliminado aleatoriamente nessa posição, e $\hat{y}_{ij(m)}$ é a m -ésima imputação nessa posição usando a equação (16) de acordo com o método proposto. Essa expressão é calculada para $l = 1, 2, \dots, na$, em que na é o número total de valores omissos. A expressão pode ser separada em dois termos,

$$acc_l = \frac{\sum_{m=1}^M (\hat{y}_{ij(m)} - \bar{Y}_l)^2}{M-1} + \frac{M(\bar{Y}_l - VO_l)^2}{M-1} \quad (18)$$

Onde \bar{Y}_l é o valor real imputado na posição l , então que o primeiro termo representa uma variância sobre os valores M em cada posição e o segundo um viés na imputação final. Assim, o primeiro termo é uma medida de precisão que refere-se aos erros aleatórios e o segundo termo é uma medida que remete ao erro sistemático na posição l .

Uma medida geral de desempenho T_{acc} pode ser calculada pela média das medidas acc_l da seguinte forma:

$$T_{acc} = \frac{\sum_{l=1}^{na} acc_l}{na} \quad (19)$$

em que $na = g \times e \times porc$, com g representando o número total de genótipos, e representando o número total de ambientes e $porc$ representando a porcentagem de dados perdidos.

É possível dividir T_{acc} em dois componentes:

$$T_{acc} = V_E + VQM \quad (20)$$

Onde,

$$V_E = \frac{1}{na} \sum_{l=1}^{na} \left[\frac{\sum_{m=1}^M (\hat{y}_{ij(m)} - \bar{Y}_l)^2}{M-1} \right] \quad (21)$$

e

$$VQM = \frac{1}{na} \sum_{l=1}^{na} \frac{M(\bar{Y}_l - VO_l)^2}{M-1} \quad (22)$$

O primeiro componente V_E representa a variação agrupada entre imputações dentro de posições, portanto, quanto maior for seu valor, menor será a precisão do método de múltiplas imputações. No entanto, um valor pequeno para este componente não significa necessariamente que o método de imputação é bom, pois o método pode ser polarizado.

O segundo componente VQM representa o viés médio quadrado entre os valores de \bar{Y} e VO , então quanto menor o viés, maior é o número de imputações que são semelhantes aos valores originais e maior é a sua precisão. Portanto, quanto menores forem os valores de V_E e VQM , melhor será o método de imputação múltipla.

Alarcón e Dias (2009), aplicaram este método em dados de algodão com o objetivo de avaliar o novo desenvolvimento usando os modelos de efeitos aditivos com interação multiplicativa (AMMI). Neste mesmo estudo comparou o método IMLD com outros procedimentos de imputação aplicados em experimentos genótipo x ambiente com dados ausentes.

Mais tarde Alarcón, Dias e Peña (2014) propuseram um novo algoritmo de imputação baseado no trabalho desenvolvido por Yan em (2003), usando as mesmas medidas de avaliação da precisão adotadas por Bergamo, Dias e Krzanowski (2008) e realizando apenas uma DVS.

Baseado no que foi descrito nesse capítulo, serão expostos, no próximo capítulo, os procedimentos utilizados nessa pesquisa.

3. MATERIAL E MÉTODOS

Nessa pesquisa foi aplicado o método de Imputação Múltipla Livre de Distribuição aos conjuntos de dados reais de um experimento balanceado desenvolvido por alunos do Programa de Pós-graduação em Agronomia do Centro de Ciências Rurais da Universidade Federal de Santa Maria.

Os bancos de dados contêm informações sobre dois experimentos realizados com o objetivo de verificar as pressuposições do modelo matemático e avaliar o efeito da aplicação de bioproduto de batata no lugar da adubação mineral na produtividade, qualidade de frutos e coloração de folhas de tomateiro. Um dos experimentos foi conduzido em túnel plástico e outro a campo, a partir do delineamento blocos ao acaso com três repetições e 12 tratamentos.

Para o presente estudo foi considerado somente os bancos de dados completos obtidos na terceira colheita conduzida em campo e na terceira colheita em túnel plástico. Nas colheitas foram avaliados o peso de frutos por planta, número de frutos por planta, comprimento e largura média de frutos.

Desta forma, foram formados 8 bancos de dados balanceados, com dados reais, cujas notações adotadas para diferenciá-los seguem no Quadro 1.

Quadro 1 – Variáveis avaliadas em cada banco de dados e notação adotada como referência para cada variável.

Experimento com tomate	Variáveis analisadas	Notação
Em campo	Peso de frutos por planta (g)	D1
	Número de frutos por planta (g)	D2
	Comprimento de frutos (mm)	D3
	Largura média de frutos(mm)	D4
Em túnel	Peso de frutos por planta (g)	D5
	Número de frutos por planta (g)	D6
	Comprimento de frutos (mm)	D7
	Largura média de frutos(mm)	D8

Após organizar os bancos de dados com as colunas sendo os blocos e as linhas os tratamentos, foram determinadas aleatoriamente as posições das observações a

serem excluídas em todos os bancos de dados balanceados. Assim, foram gerados três novos bancos de dados desbalanceados a partir de cada banco de dado completo inicial, com a exclusão de 5%, 15% e 30% das observações, porcentagens adotadas por Bergamo, Dias e Krzanowski (2008). Resultando, após arredondamento, na retirada de 1, 5 e 10 observações, respectivamente, no presente estudo.

As posições das observações retiradas foram as mesmas para todas as variáveis e o valor para os expoentes foram os mesmos utilizados no estudo de Bergamo, Dias e Krzanowski (2008).

Seguindo, foi aplicado o método de imputação IMLD e, a partir do “novo” banco de dados, agora incompleto, foram determinadas estimativas para cada dado faltante e as mesmas comparadas com o respectivo valor real que foi retirado no primeiro momento deste processo, seguindo os mesmos passos usados por Bergamo, Dias e Krzanowski (2008).

Com os bancos de dados iniciais completos, desbalanceados e completos com o método de imputação, foi dado início a análise da variabilidade dos dados e a precisão dos resultados obtidos.

Para determinar a variabilidade considerou-se o coeficiente de variação dos dados, a média de todas as observações de cada banco e de cada coluna (blocos), pois a média de cada coluna é o ponto de partida para desenvolvimento do método de imputação.

Também, verificou-se o coeficiente de variação experimental, o qual relaciona o desvio padrão em termos da porcentagem média aritmética. Em um primeiro momento é apresentado o coeficiente de variação dos valores presentes no banco de dados e em um segundo, o coeficiente de variação experimental.

Segundo Banzatto e Kronka (2006), este coeficiente é utilizado para comparar a variabilidade de seus resultados com a obtida por pesquisadores que trabalham com material semelhante e remete a uma ideia de precisão do experimento, quanto menor o coeficiente, maior a precisão.

Quando avaliado um conjunto de dados este coeficiente é dado da seguinte forma:

$$CV = \frac{s}{\hat{m}} \cdot 100 \quad (23)$$

Em que, s é o desvio padrão e \hat{m} é a estimativa da média.

Para o coeficiente de variação experimental, o desvio padrão (s) e a estimativa da média (\hat{m}) são determinadas a partir da equação que segue.

$$s = \sqrt{Q.M.Res.} \quad (24)$$

$$\hat{m} = \frac{G}{IJ} \quad (25)$$

em que,

Q. M. Res. : Quadrado médio do resíduo;

$$G = \sum_{i=1}^I \sum_{j=1}^J x_{ij}$$

x_{ij} : Valor observado no tratamento i e bloco j ;

I : número de tratamentos

J : número de blocos

O coeficiente de variação pode ser classificado conforme é apresentado no Quadro 2.

Quadro 2 – Classificação do coeficiente de variação experimental.

CV	Avaliação	Precisão
< 10%	Baixo	Alta
10% a 20%	Médio	Média
20% a 30%	Alto	Baixa
> 30%	Muito Alto	Muito Baixa

Fonte: Pimentel Gomes (1985).

Por fim, foram comparados os resultados obtidos com o objetivo de verificar se houve diferença entre os resultados dos dados com maior e menor variabilidade, se em alguns dos bancos de dados a qualidade de imputação foi melhor que em outros.

Estes procedimentos foram realizados no software R Core Team (2017) e no RStudio (2009 - 2017), utilizando o script implementado por Gandolfi (2010).

4. RESULTADOS

Ao gerar os bancos de dados incompletos e completos com 5%, 15% e 30% de dados em falta e organizá-los, foi possível obter as médias, desvio padrão e coeficiente de variação de cada coluna (blocos) de cada experimento, os quais encontram-se no ANEXO A.

Estes resultados foram considerados importantes no processo de análise devido ao fato de que, para iniciar o processo de geração dos bancos de dados imputados nas caselas vazias, é inserida a média da respectiva coluna.

As estimativas encontradas nas M=5 imputações que foram combinadas para determinar a estimativa final de cada dado retirado são apresentadas no Anexo C.

Após os primeiros passos, foi dado início a obtenção dos resultados das medidas de avaliação de precisão V_E , V_{QM} e T_{acc} . Na Tabela 1, é possível verificar estas medidas e compará-las com o resultado do coeficiente de variação do conjunto de observações dos bancos de dados iniciais.

Tabela 1 – Coeficiente de variação do conjunto de observações do banco de dados inicial e medidas de desempenho dos valores imputados.

BANCO DE DADOS	Coeficiente de Variação	MEDIDAS DE PRECISÃO								
		VE*			VQM**			Tacc***		
	Banco de dados inicial	5%	15%	30%	5%	15%	30%	5%	15%	30%
D1	51%	18	12	7	519930	170525	243185	519948	170538	243192
D2	49%	0,00007	0,00004	0,00003	0,00001	2	6	0,00008	2	6
D3	11%	0,00354	0,00003	0,00004	354	23	33	354	23	33
D4	11%	0,01709	0,00056	0,00028	778	76	45	778	76	45
D5	49%	15	6	3	16802	713611	496391	16817	713616	496395
D6	43%	0,00055	0,00005	0,00002	50	6	3	50	6	3
D7	15%	0,00342	0,00236	0,00135	65	23	60	65	23	60
D8	16%	0,00040	0,00003	0,00010	1634	286	274	1634	286	274

* VE: Variação agrupada entre imputações dentro de posições;

**VQM: Viés médio quadrado entre os valores imputados e observados;

***T_{acc}: Soma de VE e VQM.

Dos oito bancos de dados reais completos (balanceados), que foram disponibilizados para esta pesquisa, D3, D4, D7 e D8 apresentaram um baixo

coeficiente de variação enquanto D1, D2, D5 e D6 apresentam um maior coeficiente de variação.

Os bancos de dados D1 e D5 apresentaram uma maior variação agrupada entre imputações, indicada pela medida de V_E , pode-se então dizer que a precisão do método é pequena. Também se nota que quanto mais observações foram retiradas menor o valor de V_E , resultando em uma melhor precisão.

Ao avaliar esta medida em D2, D3, D4, D6, D7 e D8, percebe-se que os valores são baixos, indicando que a variação entre as imputações é baixa, portanto a precisão é boa.

Comparando os valores de V_E com os coeficientes de variação, os resultados mostram que em todos os bancos de dados que apresentaram baixo coeficiente de variação a precisão entre imputações foram boas. Porém, quando o coeficiente de variação é mais alto, alguns bancos de dados apresentaram boa precisão entre imputações e outros uma baixa precisão.

Decidiu-se então verificar o que poderia estar ocasionando estas diferenças. Para isso, avaliou-se os bancos de dados D1 que apresentou baixa precisão e coeficiente de variação alto, D2 que apresentou boa precisão e alto coeficiente de variação, e D3 que apresentou boa precisão e baixo coeficiente de variação. Foi retirado 5% das observações, de diferentes posições, isto é, na mesma coluna em que foi retirado a observação na primeira realização da imputação, foram avaliados os resultados, após a retirada de diferentes observações. Na Tabela 2, encontram-se os resultados obtidos.

Como os dados possuem uma alta variabilidade, apresentam alguns valores com grande diferença dos demais, então optou-se por retirar valores mais altos e também mais baixos para poder avaliar se isso influenciaria nos resultados.

De fato, conforme retirados diferentes valores, em alguns casos, a precisão entre as imputações mostrou-se melhor que em outros casos. Conforme pode ser observado, também, nos resultados apresentados no ANEXO B.

Conforme já mencionado, para Bergamo, Dias e Krzanowski (2008), o fato de V_E ser baixo não é garantia de que o método é bom. Ao considerar os resultados referentes ao VQM, que representa o viés médio quadrado entre os valores de \bar{Y} e $V0$, foram encontrados valores altos, o que indica que o número de imputações semelhantes aos valores originais é pequeno.

Tabela 2 – Valores de precisão dos resultados após a retirada de diferentes valores de um mesmo banco de dados.

Banco de Dados	Posição do valor retirado	Valor retirado	Valor estimado	V_E	VQM	T_{acc}
D1	(7;1)	1587	574,8	16,6204	1280678	1280694
	(8;1)	1216	917,76	5,1197	11186	111191
	(5;1)	600,5	1529,5	4,4209	1078938	1078942
	(2;1)	2054	1002	0,9085	1383265	1383266
D2	(7;1)	5,5	2,12	0,0008	14,2718	14,2726
	(8;1)	3,5	6,69	0,0006	12,7408	12,7415
	(5;1)	3	6,95	0,0007	19,5025	19,5033
	(2;1)	8	4,68	0,0001	13,0543	13,8055
D3	(7;1)	61	66,16	0,0005	33,2769	33,2774
	(8;1)	69,29	62,52	0,0001	57,3525	57,3527
	(5;1)	62	68,57	0,0022	487,1505	487,1528
	(2;1)	58,88	75,4	0,0045	341,1590	341,1635

Com objetivo de verificar se causaria alguma influência nos resultados da análise do experimento, se determinou os coeficientes de variação experimental dos bancos de dados originais, desbalanceados e balanceados após inserir as estimativas dos valores retirados, o que pode ser visto na Tabela 3.

Tabela 3 – Coeficiente de variação experimental para os bancos completo, incompleto e completo com imputação para os dados do estudo.

BANCO DE DADOS	COEFICIENTE DE VARIÇÃO						
	BANCO DE DADOS						
	COMPLETO	INCOMPLETO			COM IMPUTAÇÃO		
		5%	15%	30%	5%	15%	30%
D1	49,86%	50,24%	52,17%	47,98%	49,60%	47,20%	42,65%
D2	49,99%	51,03%	52,03%	45,10%	49,99%	47,44%	39,00%
D3	10,78%	10,71%	11,61%	13,50%	10,70%	10,45%	10,55%
D4	11,62%	11,33%	11,85%	12,06%	11,20%	11,22%	11,90%
D5	48,08%	47,45%	50,00%	51,94%	47,69%	46,61%	40,46%
D6	39,63%	39,60%	42,02%	47,07%	43,76%	39,28%	36,34%
D7	15,29%	12,67%	12,78%	10,93%	14,19%	14,44%	12,82%
D8	15,35%	12,65%	12,88%	10,87%	12,42%	12,67%	10,55%

Tomando como base a classificação do coeficiente de variação experimental apresentada no Quadro 2, temos que D1, D2, D5 e D6 apresentaram um maior coeficiente de variação, o que indica uma baixa precisão. Estes coeficientes sofreram um aumento nos dados desbalanceados e se mantiveram próximos dos valores referentes aos dados iniciais quando imputadas as estimativas para as caselas vazias.

Já em D3, D4, D7 e D8, os coeficientes experimentais encontrados são baixos e, portanto, a precisão é alta. Nos dados desbalanceados e completos com a imputação ao analisar D3 e D4 pode-se observar que os resultados se mantiveram bem próximos dos originais.

Porém, em D7 a precisão melhorou com os dados desbalanceados e os valores do banco de dados com imputação ficaram mais próximos daqueles encontrados nos dados iniciais.

Para o método aplicado, os resultados desta pesquisa indicam que nem sempre são obtidas boas aproximações para os valores reais dos dados que foram retirados.

Ao investigar o que poderia estar ocasionando isso, percebe-se que os dados onde a variabilidade é consideravelmente baixa o valor observado e o valor imputado são mais próximos. Porém, quando considerados dados com maior variabilidade nem sempre foram tão satisfatórios os resultados quanto a precisão.

Quando um valor mais alto era retirado, ao determinar a média da coluna, no ponto de partida inicial, o valor a ser imputado tende a ser menor do que o real, já quando o valor retirado era menor o valor a ser imputado aumenta. Logo, dependendo do valor que for retirado, é possível obter diferentes resultados para as medidas de precisão.

Neste trabalho, dispomos dos dados iniciais para poder comparar, mas na realidade estes valores são desconhecidos, sendo impossível saber se o dado que se encontra em falta na realidade era mais alto ou menor que os demais.

Em alguns estudos é necessário que o banco de dados seja completo para prosseguir na análise estatística, retirar todas as observações em que há uma em falta pode causar uma perda significativa de informações.

5. CONCLUSÃO

Neste estudo, a variabilidade dos dados influenciou de maneira negativa nos resultados obtidos ao aplicar o método de imputação.

Desta forma, é de grande importância que seja considerada a variabilidade dos dados em questão. Percebe-se que o fato de que os dados possuam uma baixa variabilidade é o que mais assegura resultados mais próximos dos reais.

É necessário analisar cautelosamente se, aplicar um determinado método de imputação de dados trará resultados menos viesados ou é mais confiável trabalhar com um banco de dados desbalanceados, quando a análise pode ser feita mesmo nestas condições.

No caso de um experimento conduzido no Delineamento em Blocos Casualizados, se o número de informações perdidas é relativamente baixo, talvez não seja a melhor saída recorrer para a imputação.

REFERÊNCIAS

- ALARCÓN, S. A.; 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.
- ALARCÓN, S, A.; DIAS, C. T. S.; PEÑA, M. G. **Imputação múltipla livre de distribuição em tabelas incompletas de dupla entrada**. Pesquisa agropecuária brasileira. Brasília, v.49, n.9, p. 683 – 691, 2014.
- BERGAMO, G. C.; DIAS, C. T. S.; KRZANOWSKI, W. J. **Distribution-free Multiple Imputation in na Interaction matrix through singular value decomposition**. Ciência Agrícola. Piracicaba, v.65, n.4, p. 422 – 427, 2008.
- BANZATTO, D. A.; KRONKA, S. N. **Experimentação Agrícola**. Jaboticabal, São Paulo, Funep, 4. ed., 2006.
- CARTER, R. L. **Solutions for Missing Data in Structural Equation Modeling**. Research & Praticce in Assessment, v.1, 2006.
- DIAS, A. J. R.; ALBIERI, S. **Uso de imputação em pesquisas domiciliares**. VIII Encontro Nacional de Estudos Populacionais. Anais. Volume 1: Informação Demográfica, Fecundidade, Demográfica Histórica.ABEP, São Paulo, p. 11 – 26, 1992.
- ENDERS, C. K. **A Primer on the Use of Modern Missing-Data Methods in Psychosomatic Medicine Research**, Psychosomatic Medicine, v.68, p. 427 – 436, 2006.
- ENDERS, C. K. **Applied Missing Data Analysis**. Series Editor's Note by Todd D. Little, The Guilford Press, New York, London, 2010.
- ENGELS, J.M.; DIEHR, P. **Imputation of missing longitudinal data**: a comparison of methods. Journal of Clinical Epidemiology, 2003.
- GANDOLFI, M. **Imputação Múltipla via algoritmo MICE e método IMLD**. Dissertação (mestrado), Universidade Estadual de Maringá, Maringá, 2010.
- LITTLE, R. J. A.; RUBIN, D. **Statistical Analysis with Missing Data**. Journal of Educational Statistics, v.16, n.2, p. 150 – 155, 1987
- NUNES, L. N.; KLÜCK, M. M.; FACHEL, J. M. G. **Uso da imputação múltipla de dados faltantes**: uma simulação utilizando dados epidemiológicos. Caderno de Saúde Pública, Rio de Janeiro, v.25, n.2, p. 268 – 278, 2009.
- NUNES, L. N.; KLÜCK, M. M.; FACHEL, J. M. G. **Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica**. Revista brasileira de Epidemiologia, v.13, n.4, p. 596 – 606, 2010.

PEUGH, J. L.; ENDERS, C.K. **A Review of Reporting Practices and Suggestions for Improvement.** Review of Educational Research, v.74, n.4, p. 525 – 556, 2004.

PIMENTEL GOMES, F. **Curso de Estatística Experimental.** São Paulo: Nobel, 1985.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2017.

RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA . URL <http://www.rstudio.com/>. 2009 - 2017.

RUBIN, D. B. **Inference and Missing Data.** Biometrika, v.63, n.3, p. 581 – 592, 1976.

SCHAFER, J.L.; GRAHAM, J.W. **Missing data:** Our view of the state of the art. Psychological Methods, v.7, n.2, p. 147 – 177, 2002.

ANEXO A – BANCOS DE DADOS COMPLETOS, INCOMPLETOS E COMPLETOS COM VALORES IMPUTADOS

D1 – Experimento em campo com tomate, variável peso de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1178,67	1252,88	1016,75
DESVIO PADRÃO	550,70	549,40	690,51
COEFICIENTE DE VARIAÇÃO	46,72	43,85	67,91

TRATAMENTOS	BLOCOS		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10		867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1161,55	1252,88	1016,75
DESVIO PADRÃO	574,22	549,40	690,51
COEFICIENTE DE VARIAÇÃO	49,44	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	722,06	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1124,92	1252,88	1016,75
DESVIO PADRÃO	562,01	549,40	690,51
COEFICIENTE DE VARIAÇÃO	49,96	43,85	67,91

D1 – Experimento em campo com tomate, variável peso de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1178,67	1252,88	1016,75
DESVIO PADRÃO	550,70	549,40	690,51
COEFICIENTE DE VARIAÇÃO	46,72	43,85	67,91

TRATAMENTOS	BLOCOS		
	1	2	3
1		564,5	1080,5
2	2054		1275
3	1310	944	
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	
10		867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1173,60	1252,73	1075,55
DESVIO PADRÃO	603,81	576,22	732,36
COEFICIENTE DE VARIAÇÃO	51,45	46,00	68,09

TRATAMENTOS	BLOCOS		
	1	2	3
1	1304,6	564,5	1080,5
2	2054	1553,4	1275
3	1310	944	1184,5
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	1107,3
10	1326	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1197,22	1277,78	1087,28
DESVIO PADRÃO	548,97	556,22	663,21
COEFICIENTE DE VARIAÇÃO	45,85	43,53	61,00

D1 – Experimento em campo com tomate, variável peso de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1178,67	1252,88	1016,75
DESVIO PADRÃO	550,70	549,40	690,51
COEFICIENTE DE VARIAÇÃO	46,72	43,85	67,91

TRATAMENTOS	BLOCOS		
	1	2	3
1		564,5	1080,5
2	2054		1275
3		944	
4	1950,5	1767	2928,5
5	600,5		739,5
6	291	1823	1305,5
7		239	
8	1216	1266	
9	1261	1147	
10		867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1104,88	1185,95	1151,06
DESVIO PADRÃO	660,52	560,73	786,40
COEFICIENTE DE VARIAÇÃO	59,78	47,28	68,32

TRATAMENTOS	BLOCOS		
	1	2	3
1	1154,4	564,5	1080,5
2	2054	1409,5	1275
3	1163,3	944	1243,2
4	1950,5	1767	2928,5
5	600,5	1093	739,5
6	291	1823	1305,5
7	1294,6	239	1445
8	1216	1266	1070
9	1261	1147	1099,3
10	1134,6	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1132,16	1196,83	1172,16
DESVIO PADRÃO	529,82	512,30	634,44
COEFICIENTE DE VARIAÇÃO	46,80	42,80	54,13

D2 – Experimento em campo com tomate, variável média do número de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,50	4,38	3,50
DESVIO PADRÃO	1,92	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

TRATAMENTOS	BLOCOS		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10		2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,55	4,38	3,50
DESVIO PADRÃO	2,01	2,32	1,86
COEFICIENTE DE VARIAÇÃO	44,12	52,96	53,10

TRATAMENTOS	BLOCOS		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,50	4,38	3,50
DESVIO PADRÃO	1,92	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

D2 – Experimento em campo com tomate, variável média do número de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	4.50	2.00	4.50
2	8.00	3.00	3.50
3	5.50	4.00	3.00
4	7.00	6.50	8.50
5	3.00	8.50	2.50
6	1.00	6.00	4.00
7	5.50	0.50	4.50
8	3.50	6.50	1.50
9	3.50	3.50	1.50
10	4.00	2.50	3.00
11	5.50	3.50	3.00
12	3.00	6.00	2.50
MÉDIA	4,50	4,38	3,50
DESVIO PADRÃO	1,92	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

TRATAMENTOS	BLOCOS		
	1	2	3
1			4,5
2	8		3,5
3	5,5	4	
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	
10		2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,55	4,50	3,75
DESVIO PADRÃO	2,11	2,39	1,92
COEFICIENTE DE VARIAÇÃO	46,46	53,05	51,16

TRATAMENTOS	BLOCOS		
	1	2	3
1	4,01	2	4,5
2	8	4,19	3,5
3	5,5	4	3,85
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	4,09
10	4,23	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,48	4,47	3,79
DESVIO PADRÃO	1,92	2,28	1,74
COEFICIENTE DE VARIAÇÃO	42,88	50,92	45,90

D2 – Experimento em campo com tomate, variável média do número de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,5000	4,3750	3,5000
DESVIO PADRÃO	1,9188	2,3172	1,8586
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

TRATAMENTOS	BLOCOS		
	1	2	3
1		2	4,5
2	8		3,5
3		4	
4	7	6,5	8,5
5	3		2,5
6	1	6	4
7		0,5	
8	3,5	6,5	
9	3,5	3,5	
10		2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,3125	4,1000	3,9375
DESVIO PADRÃO	2,3290	2,0923	1,9719
COEFICIENTE DE VARIAÇÃO	54,01	51,03	50,08

TRATAMENTOS	BLOCOS		
	1	2	3
1	3,79	2	4,5
2	8	3,7	3,5
3	4,21	4	3,71
4	7	6,5	8,5
5	3	3,32	2,5
6	1	6	4
7	3,44	0,5	3,6
8	3,5	6,5	4,21
9	3,5	3,5	4,11
10	4,09	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,1692	4,0017	3,9275
DESVIO PADRÃO	1,8785	1,9082	1,5808
COEFICIENTE DE VARIAÇÃO	45,06	47,68	40,25

D3 – Experimento em campo com tomate, variável média do comprimento de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,17	63,66	63,50
DESVIO PADRÃO	8,21	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,87	8,24	8,11

TRATAMENTOS	BLOCOS		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10		66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	58,70	63,66	63,50
DESVIO PADRÃO	8,43	5,24	5,15
COEFICIENTE DE VARIAÇÃO	14,37	8,24	8,11

TRATAMENTOS	BLOCOS		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	47,55	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	57,77	63,66	63,50
DESVIO PADRÃO	8,66	5,24	5,15
COEFICIENTE DE VARIAÇÃO	14,99	8,24	8,11

D3 – Experimento em campo com tomate, variável média do comprimento de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,17	63,66	63,50
DESVIO PADRÃO	8,21	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,87	8,24	8,11

TRATAMENTOS	BLOCOS		
	1	2	3
1		66,25	55,67
2	58,88		70,57
3	54	60,25	
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	
10		66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	58,93	62,93	62,99
DESVIO PADRÃO	8,86	4,82	5,51
COEFICIENTE DE VARIAÇÃO	15,03	7,66	8,75

TRATAMENTOS	BLOCOS		
	1	2	3
1	56,32	66,25	55,67
2	58,88	64,15	70,57
3	54	60,25	63,88
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	63,62
10	59,47	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	58,75	63,04	63,11
DESVIO PADRÃO	8,05	4,61	4,99
COEFICIENTE DE VARIAÇÃO	13,70	7,32	7,91

D3 – Experimento em campo com tomate, variável média do comprimento de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,17	63,66	63,50
DESVIO PADRÃO	8,21	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,87	8,24	8,11

TRATAMENTOS	BLOCOS		
	1	2	3
1		66,25	55,67
2	58,88		70,57
3		60,25	
4	66,86	62	60,65
5	48,83		66,8
6	62	66,58	66,63
7		69	
8	69,29	56,31	
9	71,71	58,43	
10		66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,28	63,73	62,46
DESVIO PADRÃO	9,83	4,25	6,10
COEFICIENTE DE VARIAÇÃO	16,58	6,66	9,77

TRATAMENTOS	BLOCOS		
	1	2	3
1	57,55	66,25	55,67
2	58,88	63,78	70,57
3	57,3	60,25	63,1
4	66,86	62	60,65
5	48,83	65,46	66,8
6	62	66,58	66,63
7	61,1	69	60,49
8	69,29	56,31	62,02
9	71,71	58,43	61,35
10	60,12	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,19	63,88	62,22
DESVIO PADRÃO	7,90	3,87	4,91
COEFICIENTE DE VARIAÇÃO	13,35	6,06	7,90

D4 – Experimento em campo com tomate, variável média da largura de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	81,11	81,75	85,22
2	84,31	101,67	93,29
3	76,45	71,5	90,83
4	85,07	83,23	92,29
5	71,17	78,18	91
6	89	87,92	92,63
7	87,09	107	78,44
8	94,14	71,38	85,67
9	94,14	90	88
10	94,63	94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	83,67	87,76	85,57
DESVIO PADRÃO	8,61	11,09	8,23
COEFICIENTE DE VARIAÇÃO	10,29	12,63	9,61

TRATAMENTOS	BLOCOS		
	1	2	3
1	81,11	81,75	85,22
2	84,31	101,67	93,29
3	76,45	71,5	90,83
4	85,07	83,23	92,29
5	71,17	78,18	91
6	89	87,92	92,63
7	87,09	107	78,44
8	94,14	71,38	85,67
9	94,14	90	88
10		94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	82,67	87,76	85,57
DESVIO PADRÃO	8,28	11,09	8,23
COEFICIENTE DE VARIAÇÃO	10,01	12,63	9,61

TRATAMENTOS	BLOCOS		
	1	2	3
1	81,11	81,75	85,22
2	84,31	101,67	93,29
3	76,45	71,5	90,83
4	85,07	83,23	92,29
5	71,17	78,18	91
6	89	87,92	92,63
7	87,09	107	78,44
8	94,14	71,38	85,67
9	94,14	90	88
10	69,68	94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	81,59	87,76	85,57
DESVIO PADRÃO	8,74	11,09	8,23
COEFICIENTE DE VARIAÇÃO	10,71	12,63	9,61

D4 – Experimento em campo com tomate, variável média da largura de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	81,11	81,75	85,22
2	84,31	101,67	93,29
3	76,45	71,5	90,83
4	85,07	83,23	92,29
5	71,17	78,18	91
6	89	87,92	92,63
7	87,09	107	78,44
8	94,14	71,38	85,67
9	94,14	90	88
10	94,63	94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	83,67	87,76	85,57
DESVIO PADRÃO	8,61	11,09	8,23
COEFICIENTE DE VARIAÇÃO	10,29	12,63	9,61

TRATAMENTOS	BLOCOS		
	1	2	3
1		81,75	85,22
2	84,31		93,29
3	76,45	71,5	
4	85,07	83,23	92,29
5	71,17	78,18	91
6	89	87,92	92,63
7	87,09	107	78,44
8	94,14	71,38	85,67
9	94,14	90	
10		94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	82,83	86,50	84,80
DESVIO PADRÃO	8,71	10,68	8,85
COEFICIENTE DE VARIAÇÃO	10,51	12,35	10,44

TRATAMENTOS	BLOCOS		
	1	2	3
1	82,14	81,75	85,22
2	84,31	89,83	93,29
3	76,45	71,5	89,76
4	85,07	83,23	92,29
5	71,17	78,18	91
6	89	87,92	92,63
7	87,09	107	78,44
8	94,14	71,38	85,67
9	94,14	90	81,31
10	83,78	94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	82,85	86,78	84,93
DESVIO PADRÃO	7,88	10,23	8,21
COEFICIENTE DE VARIAÇÃO	9,52	11,79	9,67

D4 – Experimento em campo com tomate, variável média da largura de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	81,11	81,75	85,22
2	84,31	101,67	93,29
3	76,45	71,5	90,83
4	85,07	83,23	92,29
5	71,17	78,18	91
6	89	87,92	92,63
7	87,09	107	78,44
8	94,14	71,38	85,67
9	94,14	90	88
10	94,63	94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	83,67	87,76	85,57
DESVIO PADRÃO	8,61	11,09	8,23
COEFICIENTE DE VARIAÇÃO	10,29	12,63	9,61

TRATAMENTOS	BLOCOS		
	1	2	3
1		81,75	85,22
2	84,31		93,29
3		71,5	
4	85,07	83,23	92,29
5	71,17		91
6	89	87,92	92,63
7		107	
8	94,14	71,38	
9	94,14	90	
10		94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	83,09	87,33	85,49
DESVIO PADRÃO	9,43	10,88	9,71
COEFICIENTE DE VARIAÇÃO	11,35	12,46	11,36

TRATAMENTOS	BLOCOS		
	1	2	3
1	81,09	81,75	85,22
2	84,31	87,7	93,29
3	78,15	71,5	91,22
4	85,07	83,23	92,29
5	71,17	79,39	91
6	89	87,92	92,63
7	89,07	107	80,22
8	94,14	71,38	84,51
9	94,14	90	82,18
10	83,59	94,6	66,33
11	75,73	93,86	87,67
12	71,17	92,08	75,5
MÉDIA	83,05	86,70	85,17
DESVIO PADRÃO	7,90	10,11	8,15
COEFICIENTE DE VARIAÇÃO	9,52	11,66	9,57

D5 – Experimento em túnel com tomate, variável peso de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	340,5	947,5	961
2	112	1618	1837
3	525	737,5	603
4	1767	1504	754
5	629,5	1689	1217
6	1286	1276	691,5
7	1566,5	1305	520
8	1252	663,5	1415
9	852	965,5	722,5
10	27,5	559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	779,88	1126,04	1005,67
DESVIO	594,87	371,99	402,87
COEFICIENTE DE VARIAÇÃO	76,28	33,03	40,06

TRATAMENTOS	BLOCOS		
	1	2	3
1	340,5	947,5	961
2	112	1618	1837
3	525	737,5	603
4	1767	1504	754
5	629,5	1689	1217
6	1286	1276	691,5
7	1566,5	1305	520
8	1252	663,5	1415
9	852	965,5	722,5
10		559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	848,27	1126	1005,7
DESVIO	572,28	371,99	402,87
COEFICIENTE DE VARIAÇÃO	67,464	33,035	40,06

TRATAMENTOS	BLOCOS		
	1	2	3
1	340,5	947,5	961
2	112	1618	1837
3	525	737,5	603
4	1767	1504	754
5	629,5	1689	1217
6	1286	1276	691,5
7	1566,5	1305	520
8	1252	663,5	1415
9	852	965,5	722,5
10	143,44	559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	789,54	1126	1005,7
DESVIO	582,35	371,99	402,87
COEFICIENTE DE VARIAÇÃO	73,758	33,035	40,06

D5 – Experimento em túnel com tomate, variável peso de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	340,5	947,5	961
2	112	1618	1837
3	525	737,5	603
4	1767	1504	754
5	629,5	1689	1217
6	1286	1276	691,5
7	1566,5	1305	520
8	1252	663,5	1415
9	852	965,5	722,5
10	27,5	559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	779,88	1126,04	1005,67
DESVIO	594,87	371,99	402,87
COEFICIENTE DE VARIAÇÃO	76,28	33,03	40,06

TRATAMENTOS	BLOCOS		
	1	2	3
1		947,5	961
2	112		1837
3	525	737,5	
4	1767	1504	754
5	629,5	1689	1217
6	1286	1276	691,5
7	1566,5	1305	520
8	1252	663,5	1415
9	852	965,5	
10		559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	899,05	1081,3	1074,3
DESVIO	576,52	354,7	407,7
COEFICIENTE DE VARIAÇÃO	64,126	32,802	37,952

TRATAMENTOS	BLOCOS		
	1	2	3
1	961,26	947,5	961
2	112	654,32	1837
3	525	737,5	1057,38
4	1767	1504	754
5	629,5	1689	1217
6	1286	1276	691,5
7	1566,5	1305	520
8	1252	663,5	1415
9	852	965,5	1098,23
10	1119,7	559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	922,62	1045,74	1074,84
DESVIO	525,47	359,95	368,88
COEFICIENTE DE VARIAÇÃO	56,95	34,42	34,32

D5 – Experimento em túnel com tomate, variável peso de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	340,5	947,5	961
2	112	1618	1837
3	525	737,5	603
4	1767	1504	754
5	629,5	1689	1217
6	1286	1276	691,5
7	1566,5	1305	520
8	1252	663,5	1415
9	852	965,5	722,5
10	27,5	559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	844,18	1135,86	975,91
DESVIO	594,87	371,99	402,87
COEFICIENTE DE VARIAÇÃO	70,47	32,75	41,28

TRATAMENTOS	BLOCOS		
	1	2	3
1		947,5	961
2	112		1837
3		737,5	
4	1767	1504	754
5	629,5		1217
6	1286	1276	691,5
7		1305	
8	1252	663,5	
9	852	965,5	
10		559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	862,38	1020,6	1100,9
DESVIO	584,96	307,66	390,41
COEFICIENTE DE VARIAÇÃO	67,831	30,146	35,462

TRATAMENTOS	BLOCOS		
	1	2	3
1	886,93	947,5	961
2	112	784,08	1837
3	1026,7	737,5	1058,79
4	1767	1504	754
5	629,5	979,96	1217
6	1286	1276	691,5
7	987,5	1305	1181,14
8	1252	663,5	1019,33
9	852	965,5	1095,9
10	1000,4	559,5	749,5
11	928	1229	1264,5
12	72,5	1018	1333
MÉDIA	900,05	997,46	1096,89
DESVIO	471,03	286,52	313,58
COEFICIENTE DE VARIAÇÃO	52,33	28,73	28,59

D6 – Experimento em túnel com tomate, variável média do número de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	1,5	3	3
2	1,5	5,5	7
3	2,5	5	2,5
4	7,5	5,5	3
5	2,5	6,5	4,5
6	7	6	3
7	6	5	4,5
8	6,5	4	4,5
9	3	4	3
10	1	2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	3,71	4,67	3,96
DESVIO	2,41	1,27	1,37
COEFICIENTE DE VARIAÇÃO	64,91	27,16	34,68

TRATAMENTOS	BLOCOS		
	1	2	3
1	1,5	3	3
2	1,5	5,5	7
3	2,5	5	2,5
4	7,5	5,5	3
5	2,5	6,5	4,5
6	7	6	3
7	6	5	4,5
8	6,5	4	4,5
9	3	4	3
10		2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	3,95	4,67	3,96
DESVIO	2,36	1,27	1,37
COEFICIENTE DE VARIAÇÃO	59,69	27,16	34,68

TRATAMENTOS	BLOCOS		
	1	2	3
1	1,5	3	3
2	1,5	5,5	7
3	2,5	5	2,5
4	7,5	5,5	3
5	2,5	6,5	4,5
6	7	6	3
7	6	5	4,5
8	6,5	4	4,5
9	3	4	3
10	7,33	2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	4,24	4,67	3,96
DESVIO	2,45	1,27	1,37
COEFICIENTE DE VARIAÇÃO	57,90	27,16	34,68

D6 – Experimento em túnel com tomate, variável média do número de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	1,5	3	3
2	1,5	5,5	7
3	2,5	5	2,5
4	7,5	5,5	3
5	2,5	6,5	4,5
6	7	6	3
7	6	5	4,5
8	6,5	4	4,5
9	3	4	3
10	1	2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	3,71	4,67	3,96
DESVIO	2,41	1,27	1,37
COEFICIENTE DE VARIAÇÃO	64,91	27,16	34,68

TRATAMENTOS	BLOCOS		
	1	2	3
1		3	3
2	1,5		7
3	2,5	5	
4	7,5	5,5	3
5	2,5	6,5	4,5
6	7	6	3
7	6	5	4,5
8	6,5	4	4,5
9	3	4	
10		2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	4,20	4,59	4,20
DESVIO	2,34	1,30	1,38
COEFICIENTE DE VARIAÇÃO	55,61	28,32	32,82

TRATAMENTOS	BLOCOS		
	1	2	3
1	3,91	3	3
2	1,5	3,29	7
3	2,5	5	4,52
4	7,5	5,5	3
5	2,5	6,5	4,5
6	7	6	3
7	6	5	4,5
8	6,5	4	4,5
9	3	4	4,32
10	3,79	2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	4,14	4,48	4,24
DESVIO	2,12	1,30	1,25
COEFICIENTE DE VARIAÇÃO	51,12	28,90	29,52

D6 – Experimento em túnel com tomate, variável média do número de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	1,5	3	3
2	1,5	5,5	7
3	2,5	5	2,5
4	7,5	5,5	3
5	2,5	6,5	4,5
6	7	6	3
7	6	5	4,5
8	6,5	4	4,5
9	3	4	3
10	1	2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	3,71	4,67	3,96
DESVIO	2,41	1,27	1,37
COEFICIENTE DE VARIAÇÃO	64,91	27,16	34,68

TRATAMENTOS	BLOCOS		
	1	2	3
1		3	3
2	1,5		7
3		5	
4	7,5	5,5	3
5	2,5		4,5
6	7	6	3
7		5	
8	6,5	4	
9	3	4	
10		2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	4,19	4,40	4,13
DESVIO	2,48	1,20	1,55
COEFICIENTE DE VARIAÇÃO	59,17	27,21	37,64

TRATAMENTOS	BLOCOS		
	1	2	3
1	3,33	3	3
2	1,5	3,89	7
3	3,96	5	4,43
4	7,5	5,5	3
5	2,5	4,63	4,5
6	7	6	3
7	4,58	5	4,34
8	6,5	4	4,27
9	3	4	4,21
10	2,9	2	2,5
11	4	5	4,5
12	1,5	4,5	5,5
MÉDIA	4,02	4,38	4,19
DESVIO	2,03	1,10	1,24
COEFICIENTE DE VARIAÇÃO	50,42	25,04	29,68

D7 – Experimento em túnel com tomate, variável média do comprimento de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	51	68,17	67
2	39,33	63,27	64
3	66,4	50,5	63,4
4	58,6	64,45	64,5
5	64,4	62,23	66,33
6	57,29	57	62,5
7	61,75	60,8	46,33
8	59,85	53,75	68,67
9	65,67	64,13	63,83
10	35	69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	54,81	61,19	63,70
DESVIO	11,66	5,50	6,92
COEFICIENTE DE VARIAÇÃO	21,28	8,99	10,86

TRATAMENTOS	BLOCOS		
	1	2	3
1	51	68,17	67
2	39,33	63,27	64
3	66,4	50,5	63,4
4	58,6	64,45	64,5
5	64,4	62,23	66,33
6	57,29	57	62,5
7	61,75	60,8	46,33
8	59,85	53,75	68,67
9	65,67	64,13	63,83
10		69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	56,61	61,19	63,70
DESVIO	10,33	5,50	6,92
COEFICIENTE DE VARIAÇÃO	18,26	8,99	10,86

TRATAMENTOS	BLOCOS		
	1	2	3
1	51	68,17	67
2	39,33	63,27	64
3	66,4	50,5	63,4
4	58,6	64,45	64,5
5	64,4	62,23	66,33
6	57,29	57	62,5
7	61,75	60,8	46,33
8	59,85	53,75	68,67
9	65,67	64,13	63,83
10	42,2	69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	55,41	61,19	63,70
DESVIO	10,70	5,50	6,92
COEFICIENTE DE VARIAÇÃO	19,30	8,99	10,86

D7 – Experimento em túnel com tomate, variável média do comprimento de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	51	68,17	67
2	39,33	63,27	64
3	66,4	50,5	63,4
4	58,6	64,45	64,5
5	64,4	62,23	66,33
6	57,29	57	62,5
7	61,75	60,8	46,33
8	59,85	53,75	68,67
9	65,67	64,13	63,83
10	35	69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	54,81	61,19	63,70
DESVIO	11,66	5,50	6,92
COEFICIENTE DE VARIAÇÃO	21,28	8,99	10,86

TRATAMENTOS	BLOCOS		
	1	2	3
1		68,17	67
2	39,33		64
3	66,4	50,5	
4	58,6	64,45	64,5
5	64,4	62,23	66,33
6	57,29	57	62,5
7	61,75	60,8	46,33
8	59,85	53,75	68,67
9	65,67	64,13	
10		69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	57,17	61,00	63,71
DESVIO	10,72	5,73	7,65
COEFICIENTE DE VARIAÇÃO	18,75	9,39	12,00

TRATAMENTOS	BLOCOS		
	1	2	3
1	49,49	68,17	67
2	39,33	63,31	64
3	66,4	50,5	62,51
4	58,6	64,45	64,5
5	64,4	62,23	66,33
6	57,29	57	62,5
7	61,75	60,8	46,33
8	59,85	53,75	68,67
9	65,67	64,13	59,17
10	43,14	69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	55,36	61,19	63,23
DESVIO	10,66	5,50	7,04
COEFICIENTE DE VARIAÇÃO	19,26	8,99	11,14

D7 – Experimento em túnel com tomate, variável média do comprimento de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	51	68,17	67
2	39,33	63,27	64
3	66,4	50,5	63,4
4	58,6	64,45	64,5
5	64,4	62,23	66,33
6	57,29	57	62,5
7	61,75	60,8	46,33
8	59,85	53,75	68,67
9	65,67	64,13	63,83
10	35	69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	54,81	61,19	63,70
DESVIO	11,66	5,50	6,92
COEFICIENTE DE VARIAÇÃO	21,28	8,99	10,86

TRATAMENTOS	BLOCOS		
	1	2	3
1		68,17	67
2	39,33		64
3		50,5	
4	58,6	64,45	64,5
5	64,4		66,33
6	57,29	57	62,5
7		60,8	
8	59,85	53,75	
9	65,67	64,13	
10		69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	55,44	60,88	65,27
DESVIO	11,36	6,02	5,07
COEFICIENTE DE VARIAÇÃO	20,49	9,89	7,77

TRATAMENTOS	BLOCOS		
	1	2	3
1	48,48	68,17	67
2	39,33	63,36	64
3	64,07	50,5	63,12
4	58,6	64,45	64,5
5	64,4	61,06	66,33
6	57,29	57	62,5
7	65,69	60,8	65,58
8	59,85	53,75	63,88
9	65,67	64,13	63,03
10	42,59	69,25	73,8
11	62,38	58,7	68
12	36	62	56
MÉDIA	55,36	61,10	64,81
DESVIO	10,86	5,49	4,15
COEFICIENTE DE VARIAÇÃO	19,61	8,99	6,40

D8 – Experimento em túnel com tomate, variável média da largura de frutos:

- 5% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	79	89,83	93,33
2	50,33	87,45	83,29
3	78,8	67,6	77,8
4	77,93	86,27	80,83
5	80,4	79,46	84
6	67,36	70	77,33
7	81,33	81,2	59,11
8	72,23	67,63	89,44
9	88,33	81,5	81,33
10	38,5	85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	69,84	79,46	81,57
DESVIO	16,02	7,65	8,53
COEFICIENTE DE VARIAÇÃO	22,94	9,62	10,46

TRATAMENTOS	BLOCOS		
	1	2	3
1	79	89,83	93,33
2	50,33	87,45	83,29
3	78,8	67,6	77,8
4	77,93	86,27	80,83
5	80,4	79,46	84
6	67,36	70	77,33
7	81,33	81,2	59,11
8	72,23	67,63	89,44
9	88,33	81,5	81,33
10		85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	72,69	79,46	81,57
DESVIO	13,23	7,65	8,53
COEFICIENTE DE VARIAÇÃO	18,21	9,62	10,46

TRATAMENTOS	BLOCOS		
	1	2	3
1	79	89,83	93,33
2	50,33	87,45	83,29
3	78,8	67,6	77,8
4	77,93	86,27	80,83
5	80,4	79,46	84
6	67,36	70	77,33
7	81,33	81,2	59,11
8	72,23	67,63	89,44
9	88,33	81,5	81,33
10		74,65	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	72,85	79,46	81,57
DESVIO	12,63	7,65	8,53
COEFICIENTE DE VARIAÇÃO	17,34	9,62	10,46

D8 – Experimento em túnel com tomate, variável média da largura de frutos:

- 15% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	79	89,83	93,33
2	50,33	87,45	83,29
3	78,8	67,6	77,8
4	77,93	86,27	80,83
5	80,4	79,46	84
6	67,36	70	77,33
7	81,33	81,2	59,11
8	72,23	67,63	89,44
9	88,33	81,5	81,33
10	38,5	85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	69,84	79,46	81,57
DESVIO	16,02	7,65	8,53
COEFICIENTE DE VARIÇÃO	22,94	9,62	10,46

TRATAMENTOS	BLOCOS		
	1	2	3
1		89,83	93,33
2	50,33		83,29
3	78,8	67,6	
4	77,93	86,27	80,83
5	80,4	79,46	84
6	67,36	70	77,33
7	81,33	81,2	59,11
8	72,23	67,63	89,44
9	88,33	81,5	
10		85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	72,06	78,74	81,97
DESVIO	13,77	7,57	9,33
COEFICIENTE DE VARIÇÃO	19,11	9,62	11,39

TRATAMENTOS	BLOCOS		
	1	2	3
1	68,45	89,83	93,33
2	50,33	80,51	83,29
3	78,8	67,6	79,45
4	77,93	86,27	80,83
5	80,4	79,46	84
6	67,36	70	77,33
7	81,33	81,2	59,11
8	72,23	67,63	89,44
9	88,33	81,5	80,33
10	69,79	85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	71,57	78,89	81,62
DESVIO	12,51	7,24	8,48
COEFICIENTE DE VARIÇÃO	17,49	9,18	10,39

D8 – Experimento em túnel com tomate, variável média da largura de frutos:

- 30% dos dados retirados

TRATAMENTOS	BLOCOS		
	1	2	3
1	79	89,83	93,33
2	50,33	87,45	83,29
3	78,8	67,6	77,8
4	77,93	86,27	80,83
5	80,4	79,46	84
6	67,36	70	77,33
7	81,33	81,2	59,11
8	72,23	67,63	89,44
9	88,33	81,5	81,33
10	38,5	85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	69,84	79,46	81,57
DESVIO	16,02	7,65	8,53
COEFICIENTE DE VARIÇÃO	22,94	9,62	10,46

TRATAMENTOS	BLOCOS		
	1	2	3
1		89,83	93,33
2	50,33		83,29
3		67,6	
4	77,93	86,27	80,83
5	80,4		84
6	67,36	70	77,33
7		81,2	
8	72,23	67,63	
9	88,33	81,5	
10		85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	70,06	78,67	83,89
DESVIO	14,85	7,98	5,02
COEFICIENTE DE VARIÇÃO	21,20	10,14	5,98

TRATAMENTOS	BLOCOS		
	1	2	3
1	65,52	89,83	93,33
2	50,33	80,87	83,29
3	75,89	67,6	83,29
4	77,93	86,27	80,83
5	80,4	76,19	84
6	67,36	70	77,33
7	64,27	81,2	84,35
8	72,23	67,63	83,11
9	88,33	81,5	82,68
10	69,25	85,75	87,6
11	77,88	80,2	85,22
12	46	76,67	79,55
MÉDIA	69,62	78,64	83,72
DESVIO	12,17	7,28	4,03
COEFICIENTE DE VARIÇÃO	17,49	9,26	4,81

ANEXO B – VALORES DE PRECISÃO DOS RESULTADOS, RETIRANDO DIFERENTES VALORES DE UM MESMO BANCO DE DADOS

D1 – Observação da posição (7;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1178,67	1252,88	1016,75
DESVIO PADRÃO	550,70	549,40	690,51
COEFICIENTE DE VARIAÇÃO	46,72	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7		239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1141,55	1252,88	1016,75
DESVIO PADRÃO	561,61	549,40	690,51
COEFICIENTE DE VARIAÇÃO	49,20	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	574,8	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1094,32	1252,88	1016,75
DESVIO PADRÃO	559,91	549,40	690,51
COEFICIENTE DE VARIAÇÃO	51,17	43,85	67,91

$$V_E = 16,6204$$

$$V_{QM} = 1280678$$

$$T_{acc} = 1280694$$

D1 – Observação da posição (8;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1178,67	1252,88	1016,75
DESVIO PADRÃO	550,70	549,40	690,51
COEFICIENTE DE VARIAÇÃO	46,72	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8		1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1175,27	1252,88	1016,75
DESVIO PADRÃO	577,45	549,40	690,51
COEFICIENTE DE VARIAÇÃO	49,13	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	917,76	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1153,81	1252,88	1016,75
DESVIO PADRÃO	555,57	549,40	690,51
COEFICIENTE DE VARIAÇÃO	48,15	43,85	67,91

$$V_E = 5,1197$$

$$V_{QM} = 11186$$

$$T_{acc} = 111191$$

D1 – Observação da posição (5;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1178,67	1252,88	1016,75
DESVIO PADRÃO	550,70	549,40	690,51
COEFICIENTE DE VARIAÇÃO	46,72	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5		1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1231,23	1252,88	1016,75
DESVIO PADRÃO	545,10	549,40	690,51
COEFICIENTE DE VARIAÇÃO	44,27	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	1529,6	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1256,09	1252,88	1016,75
DESVIO PADRÃO	526,82	549,40	690,51
COEFICIENTE DE VARIAÇÃO	41,94	43,85	67,91

$$V_E = 4,4209$$

$$V_{QM} = 1078938$$

$$T_{acc} = 1078942$$

D1 – Observação da posição (2;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	2054	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1178,67	1252,88	1016,75
DESVIO PADRÃO	550,70	549,40	690,51
COEFICIENTE DE VARIAÇÃO	46,72	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2		1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1099,09	1252,88	1016,75
DESVIO PADRÃO	500,01	549,40	690,51
COEFICIENTE DE VARIAÇÃO	45,49	43,85	67,91

TRATAMENTOS	BLOCO		
	1	2	3
1	1041	564,5	1080,5
2	1002	1254,5	1275
3	1310	944	1047
4	1950,5	1767	2928,5
5	600,5	1920,5	739,5
6	291	1823	1305,5
7	1587	239	1141
8	1216	1266	406
9	1261	1147	398,5
10	1367	867,5	495
11	1055,5	1262	921
12	410,5	1979,5	463,5
MÉDIA	1091,00	1252,88	1016,75
DESVIO PADRÃO	477,57	549,40	690,51
COEFICIENTE DE VARIAÇÃO	43,77	43,85	67,91

$$V_E = 0,9085$$

$$V_{QM} = 1383265$$

$$T_{acc} = 1383266$$

D2 – Observação da posição (7;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,50	4,38	3,50
DESVIO PADRÃO	1,92	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	0,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,41	4,38	3,50
DESVIO PADRÃO	1,99	2,32	1,86
COEFICIENTE DE VARIAÇÃO	45,02	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	2,12	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,22	4,38	3,50
DESVIO PADRÃO	2,00	2,32	1,86
COEFICIENTE DE VARIAÇÃO	47,53	52,96	53,10

$$V_E = 0,0008$$

$$VQM = 14,2718$$

$$T_{acc} = 14,2726$$

D2 – Observação da posição (8;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,50	4,38	3,50
DESVIO PADRÃO	1,92	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	0,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,59	4,38	3,50
DESVIO PADRÃO	1,99	2,32	1,86
COEFICIENTE DE VARIAÇÃO	43,24	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	6,69	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,77	4,38	3,50
DESVIO PADRÃO	1,99	2,32	1,86
COEFICIENTE DE VARIAÇÃO	41,70	52,96	53,10

$$V_E = 0,0006$$

$$VQM = 12,7408$$

$$T_{acc} = 12,74145$$

D2 – Observação da posição (5;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,50	4,38	3,50
DESVIO PADRÃO	1,92	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5		8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,64	4,38	3,50
DESVIO PADRÃO	1,95	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,07	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	6,95	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,83	4,38	3,50
DESVIO PADRÃO	1,98	2,32	1,86
COEFICIENTE DE VARIAÇÃO	40,92	52,96	53,10

$$V_E = 0,0007$$

$$VQM = 19,5025$$

$$T_{acc} = 19,5033$$

D2 – Observação da posição (2;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	8	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,50	4,38	3,50
DESVIO PADRÃO	1,92	2,32	1,86
COEFICIENTE DE VARIAÇÃO	42,64	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2		3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,18	4,38	3,50
DESVIO PADRÃO	1,65	2,32	1,86
COEFICIENTE DE VARIAÇÃO	39,39	52,96	53,10

TRATAMENTOS	BLOCO		
	1	2	3
1	4,5	2	4,5
2	4,68	3	3,5
3	5,5	4	3
4	7	6,5	8,5
5	3	8,5	2,5
6	1	6	4
7	5,5	0,5	4,5
8	3,5	6,5	1,5
9	3,5	3,5	1,5
10	4	2,5	3
11	5,5	3,5	3
12	3	6	2,5
MÉDIA	4,22	4,38	3,50
DESVIO PADRÃO	1,58	2,32	1,86
COEFICIENTE DE VARIAÇÃO	37,35	52,96	53,10

$$V_E = 0,0001$$

$$VQM = 13,0543$$

$$T_{acc} = 13,8055$$

D3 – Observação da posição (7;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,17	63,66	63,50
DESVIO PADRÃO	8,21	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,87	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7		69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,01	63,66	63,50
DESVIO PADRÃO	8,59	5,24	5,15
COEFICIENTE DE VARIAÇÃO	14,55	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	66,16	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,60	63,66	63,50
DESVIO PADRÃO	8,44	5,24	5,15
COEFICIENTE DE VARIAÇÃO	14,17	8,24	8,11

$$V_E = 0,0005$$

$$V_{QM} = 33,2769$$

$$T_{acc} = 33,2774$$

D3 – Observação da posição (8;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,17	63,66	63,50
DESVIO PADRÃO	8,21	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,87	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8		56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	58,25	63,66	63,50
DESVIO PADRÃO	7,93	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,62	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	62,52	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	58,61	63,66	63,50
DESVIO PADRÃO	7,66	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,08	8,24	8,11

$$V_E = 0,0001$$

$$V_{QM} = 57,3525$$

$$T_{acc} = 57,3527$$

D3 – Observação da posição (5;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,17	63,66	63,50
DESVIO PADRÃO	8,21	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,87	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5		54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	60,11	63,66	63,50
DESVIO PADRÃO	7,90	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,14	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	68,57	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	60,82	63,66	63,50
DESVIO PADRÃO	7,92	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,02	8,24	8,11

$$V_E = 0,0022$$

$$VQM = 487,1505$$

$$T_{acc} = 487,1528$$

D3 – Observação da posição (2;1) excluída:

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	58,88	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,17	63,66	63,50
DESVIO PADRÃO	8,21	5,24	5,15
COEFICIENTE DE VARIAÇÃO	13,87	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2		71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	59,20	63,66	63,50
DESVIO PADRÃO	8,61	5,24	5,15
COEFICIENTE DE VARIAÇÃO	14,54	8,24	8,11

TRATAMENTOS	BLOCO		
	1	2	3
1	56,44	66,25	55,67
2	75,4	71,67	70,57
3	54	60,25	67,17
4	66,86	62	60,65
5	48,83	54,94	66,8
6	62	66,58	66,63
7	61	69	64,22
8	69,29	56,31	66
9	71,71	58,43	65
10	64,38	66,8	52,83
11	50,36	67,14	66
12	46,33	64,58	60,5
MÉDIA	60,55	63,66	63,50
DESVIO PADRÃO	9,45	5,24	5,15
COEFICIENTE DE VARIAÇÃO	15,60	8,24	8,11

$$V_E = 0,0045$$

$$VQM = 341,1590$$

$$T_{acc} = 341,1635$$

ANEXO C – ANÁLISE DAS 5 IMPUTAÇÕES PARA CADA DADO AUSENTE

D1

Posição da Observação Excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	1367	716,7343	719,4146	722,0788	724,7269	727,359	722,0627	17,63837395	0,005816393	519947,7572
15% de dados retirados										
(1;1)	1041	1304,7591	1304,701	1304,6439	1304,5878	1304,5327	1304,6449	0,008013021	6,8599E-05	86885,79963
(10;1)	1367	1323,1542	1323,0969	1323,0408	1322,9857	1322,9317	1323,0418	0,00773895	6,64747E-05	2415,405325
(2;2)	1254,5	1563,2706	1558,2643	1553,3348	1548,481	1543,7015	1553,4104	59,83792299	0,004979683	111744,1518
(3;3)	1047	1184,2421	1184,3813	1184,5143	1184,6414	1184,7627	1184,5084	0,0423693	0,000173767	23635,7287
(9;3)	398,5	1108,8277	1108,0474	1107,2862	1106,5438	1105,8199	1107,305	1,413805077	0,001073796	628007,0738
30% de dados retirados										
(1;1)	1041	1153,9792	1154,2059	1154,4337	1154,6627	1154,8928	1154,4349	0,130424972	0,000312823	16084,46475
(3;1)	1310	1162,9254	1163,1231	1163,3204	1163,5174	1163,714	1163,3201	0,097187004	0,000267959	26893,85317
(7;1)	1587	1293,3428	1293,9711	1294,5985	1295,2249	1295,8503	1294,5975	0,982427646	0,000765632	106874,9953
(10;1)	1367	1133,6868	1134,1523	1134,6153	1135,076	1135,5342	1134,6129	0,533275302	0,000643614	67505,22696
(2;2)	1254,5	1419,2717	1414,2828	1409,373	1404,5408	1399,785	1409,4507	59,33521907	0,005465196	30071,46883
(5;2)	1920,5	1089,9079	1091,481	1093,0415	1094,5897	1096,1257	1093,0292	6,040655611	0,00224861	855891,0296
(3;3)	1047	1242,6751	1242,9369	1243,1923	1243,4415	1243,6845	1243,1861	0,159219417	0,000320972	48111,3719
(7;3)	1141	1442,9832	1443,999	1444,991	1445,9594	1446,9049	1444,9675	2,403354956	0,001072877	115497,7047
(8;3)	406	1071,1339	1070,5652	1070,0152	1069,4836	1068,9702	1070,0336	0,731736344	0,000799425	551176,5423
(9;3)	398,5	1100,1015	1099,6603	1099,2352	1098,8261	1098,4324	1099,2511	0,435422363	0,000600301	613815,5656

D2

Posição da observação excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	4	3,9871	3,9924	3,9977	4,0029	4,0081	3,9977	6,89E-05	0,002076496	7,587E-05
15% de dados retirados										
(1;1)	4,5	4	4,0058	4,0116	4,0173	4,023	4,0115	8,28E-05	0,002266373	0,298324123
(10;1)	4	4,224	4,2262	4,2283	4,2305	4,2326	4,2283	1,15E-05	0,000803998	0,065174085
(2;2)	3	4,1817	4,1881	4,1945	4,2008	4,207	4,1944	9,97E-05	0,002386229	1,783399098
(3;3)	3	3,8481	3,8496	3,851	3,8524	3,8538	3,851	5,04E-06	0,000583083	0,905213743
(9;3)	1,5	4,0851	4,0855	4,0858	4,0862	4,0866	4,0858	3,47E-07	0,000143339	8,358210975
30% de dados retirados										
(1;1)	4,5	3,7842	3,7869	3,7896	3,7923	3,7949	3,7896	1,77E-05	0,001118216	0,630888678
(3;1)	5,5	4,2108	4,2104	4,2101	4,2098	4,2095	4,2101	2,38E-07	0,000120413	2,079738275
(7;1)	5,5	3,4292	3,4329	3,4365	3,4402	3,4438	3,4365	3,31E-05	0,00167938	5,322470445
(10;1)	4	4,0918	4,093	4,0941	4,0953	4,0964	4,0941	3,26E-06	0,000444177	0,011076525
(2;2)	3	3,6787	3,6883	3,6978	3,7072	3,7164	3,6977	0,000222744	0,004032429	0,608669055
(5;2)	8,5	3,3096	3,3141	3,3186	3,323	3,3274	3,3185	4,99E-05	0,002120267	33,55945917
(3;3)	3	3,7043	3,7065	3,7086	3,7107	3,7128	3,7086	1,13E-05	0,000903894	0,627618258
(7;3)	4,5	3,5974	3,5975	3,5975	3,5976	3,5976	3,5975	6,30E-09	2,32566E-05	1,018087695
(8;3)	1,5	4,2096	4,2097	4,2097	4,2097	4,2098	4,2097	3,57E-09	1,67971E-05	9,178092618
(9;3)	1,5	4,1073	4,1064	4,1055	4,1045	4,1037	4,1055	2,08E-06	0,000350615	8,48565961

D3

Posição da observação excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	64,38	47,5786	47,5667	47,5548	47,5429	47,531	47,5548	0,000354183	0,00039566	353,8595478
15% de dados retirados										
(1;1)	56,44	56,3171	56,3182	56,3192	56,3203	56,3214	56,3192	2,87E-06	3,00437E-05	0,018231585
(10;1)	64,38	59,4729	59,4725	59,4721	59,4717	59,4713	59,4721	3,93E-07	1,06345E-05	30,10935341
(2;2)	71,67	64,1655	64,1589	64,1523	64,1458	64,1392	64,1523	0,000107654	0,000161929	70,64412276
(3;3)	67,17	63,878	63,881	63,8839	63,8869	63,8898	63,8839	2,19E-05	7,30143E-05	13,49792397
(9;3)	65	63,6217	63,6191	63,6166	63,6141	63,6116	63,6166	1,59E-05	6,26345E-05	2,392191158
30% de dados retirados										
(1;1)	56,44	57,5456	57,5454	57,5452	57,545	57,5449	57,5452	7,96E-08	4,9762E-06	1,526889143
(3;1)	54	57,2885	57,2937	57,2989	57,3041	57,3092	57,2989	6,74E-05	0,000142941	13,60332865
(7;1)	61	61,1116	61,1061	61,1006	61,0951	61,0896	61,1006	7,57E-05	0,000142327	0,012726075
(10;1)	64,38	60,1296	60,1252	60,1208	60,1164	60,112	60,1208	4,85E-05	0,000115717	22,6760292
(2;2)	71,67	63,7931	63,7883	63,7835	63,7788	63,7741	63,7836	5,67E-05	0,00011775	77,74497625
(5;2)	54,94	65,4657	65,4624	65,4592	65,4559	65,4527	65,4592	2,62E-05	7,85035E-05	138,3164612
(3;3)	67,17	63,1008	63,1013	63,1017	63,1021	63,1025	63,1017	4,33E-07	1,05359E-05	20,68903497
(7;3)	64,22	60,4852	60,49	60,4948	60,4995	60,5043	60,4948	5,67E-05	0,000124673	17,34682321
(8;3)	66	62,0255	62,0215	62,0176	62,0137	62,0098	62,0176	3,85E-05	9,99417E-05	19,8242265
(9;3)	65	61,361	61,3579	61,3548	61,3517	61,3486	61,3548	2,43E-05	7,98883E-05	16,60937783

D4

Posição da observação excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	94,63	69,5141	69,5982	69,6814	69,7636	69,8449	69,6804	0,017091296	0,001876612	778,1177792
15% de dados retirados										
(1;1)	81,11	82,1258	82,1319	82,138	82,1441	82,1501	82,138	9,27E-05	0,000117039	1,321021018
(10;1)	94,63	83,7983	83,7887	83,779	83,7693	83,7595	83,7789	0,000234941	0,000183067	147,1815716
(2;2)	101,67	89,8835	89,8569	89,8304	89,8041	89,778	89,8306	0,0017399	0,000464327	175,2165722
(3;3)	90,83	89,7874	89,7738	89,7604	89,747	89,7337	89,7605	0,000449456	0,000236396	1,430345012
(9;3)	88	81,2836	81,2947	81,3058	81,3169	81,3279	81,3058	0,000306537	0,000215471	56,01603368
30% de dados retirados										
(1;1)	81,11	81,0715	81,0805	81,0894	81,0982	81,107	81,0893	0,000196624	0,000172956	0,000731275
(3;1)	76,45	78,1163	78,1332	78,1499	78,1665	78,183	78,1498	0,00069462	0,000337274	3,612259797
(7;1)	87,09	89,0955	89,0819	89,0684	89,0551	89,042	89,0686	0,000446973	0,000237528	4,893921107
(10;1)	94,63	83,582	83,5879	83,5938	83,5995	83,6052	83,5937	8,41E-05	0,000109709	152,250533
(2;2)	101,67	87,7138	87,7081	87,7025	87,697	87,6915	87,7026	7,76E-05	0,000100421	243,8611044
(5;2)	78,18	79,3568	79,3731	79,3893	79,4054	79,4213	79,3892	0,000650216	0,000321254	1,828295797
(3;3)	90,83	91,2405	91,2301	91,2197	91,2094	91,1992	91,2198	0,000267307	0,000179054	0,190177338
(7;3)	78,44	80,2047	80,2136	80,2224	80,2312	80,2398	80,2223	0,000193486	0,000173052	3,971112573
(8;3)	85,67	84,5033	84,505	84,5067	84,5084	84,51	84,5067	6,94E-06	3,14354E-05	1,691648835
(9;3)	88	82,1615	82,1693	82,1771	82,1848	82,1925	82,177	0,000150323	0,000149116	42,38372911

D5

Posição da observação excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	27,5	138,5819	141,0204	143,4483	145,8658	148,2729	143,4379	14,67428526	0,026706344	16816,6585
15% de dados retirados										
(1;1)	340,5	960,7568	961,0114	961,2629	961,5114	961,7568	961,2599	0,1562619	0,000411226	481678,661
(10;1)	27,5	1119,0226	1119,3404	1119,6543	1119,9643	1120,2705	1119,6504	0,243342934	0,000440567	1490990,918
(2;2)	1618	648,2188	651,3002	654,3527	657,3766	660,3721	654,3241	23,07879422	0,007342002	1160862,177
(3;3)	603	1059,6184	1058,4895	1057,3724	1056,2668	1055,1727	1057,384	3,088088163	0,001661957	258084,0671
(9;3)	722,5	1099,6077	1098,9116	1098,2222	1097,5394	1096,8631	1098,2288	1,177040042	0,000987864	176466,3409
30% de dados retirados										
(1;1)	340,5	886,0586	886,4974	886,9323	887,3633	887,7906	886,9284	0,468737544	0,000771908	373230,5188
(3;1)	525	1027,719	1027,196	1026,6792	1026,1685	1025,6638	1026,6853	0,659982138	0,000791278	314610,8353
(7;1)	1566,5	991,1966	989,3286	987,4785	985,6461	983,8312	987,4962	8,47669767	0,002948311	419065,227
(10;1)	27,5	999,7602	1000,1038	1000,4456	1000,7855	1001,1235	1000,4437	0,290405681	0,000538663	1183274,643
(2;2)	1618	778,5873	781,3627	784,1082	786,824	789,5106	784,0786	18,64408903	0,005506906	869299,8539
(5;2)	1689	978,3009	979,1394	979,9678	980,7864	981,5951	979,9579	1,695656546	0,001328796	628427,5346
(3;3)	603	1060,5529	1059,6599	1058,7782	1057,9075	1057,0477	1058,7892	1,919823651	0,001308626	259681,7089
(7;3)	520	1181,6691	1181,3998	1181,1337	1180,8708	1180,611	1181,1369	0,174928372	0,000354111	546377,6426
(8;3)	1415	1017,8614	1018,6069	1019,3424	1020,068	1020,7839	1019,3325	1,334541674	0,001133323	195692,278
(9;3)	722,5	1096,7975	1096,3434	1095,895	1095,452	1095,0146	1095,9005	0,496681809	0,000643091	174285,4134

D6

Posição da observação excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	1	7,3035	7,3183	7,3331	7,3479	7,3626	7,3331	0,000546812	0,003186826	50,13542398
15% de dados retirados										
(1;1)	1,5	3,9035	3,9066	3,9098	3,9129	3,9159	3,9097	2,43E-05	0,001257787	7,258582768
(10;1)	1	3,7753	3,7824	3,7895	3,7964	3,8034	3,7894	0,000123422	0,002929164	9,726063655
(2;2)	5,5	3,2774	3,283	3,2885	3,2939	3,2993	3,2884	7,45E-05	0,00263017	6,113932428
(3;3)	2,5	4,5212	4,5195	4,5177	4,516	4,5143	4,5177	7,60E-06	0,000605503	5,089100868
(9;3)	3	4,3282	4,3257	4,3232	4,3208	4,3183	4,3233	1,54E-05	0,000903375	2,188720375
30% de dados retirados										
(1;1)	1,5	3,3264	3,3276	3,3288	3,33	3,3311	3,3288	3,54E-06	0,000560569	4,180548843
(3;1)	2,5	3,9618	3,9624	3,9631	3,9638	3,9645	3,9631	1,14E-06	0,000271412	2,675901325
(7;1)	6	4,5765	4,5794	4,5823	4,5851	4,5879	4,5823	2,03E-05	0,000983458	2,51257458
(10;1)	1	2,894	2,8976	2,9011	2,9046	2,908	2,901	3,07E-05	0,001907667	4,517567033
(2;2)	5,5	3,8844	3,8888	3,8931	3,8973	3,9015	3,893	4,56E-05	0,001734339	3,228026488
(5;2)	6,5	4,6335	4,6337	4,6339	4,6341	4,6343	4,6339	9,61E-08	6,82423E-05	4,352911613
(3;3)	2,5	4,4357	4,4333	4,4308	4,4284	4,426	4,4308	1,48E-05	0,000867164	4,660193645
(7;3)	4,5	4,3437	4,3414	4,3392	4,337	4,3349	4,3392	1,21E-05	0,000801739	0,032316825
(8;3)	4,5	4,2787	4,2743	4,27	4,2657	4,2615	4,27	4,60E-05	0,001592284	0,06614823
(9;3)	3	4,2192	4,2164	4,2136	4,2109	4,2082	4,2136	1,89E-05	0,001031961	1,841232153

D7

Posição da observação excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	35	42,1269	42,1641	42,2011	42,2381	42,2749	42,201	0,003422847	0,001386275	64,82178381
15% de dados retirados										
(1;1)	51	49,4356	49,4643	49,4928	49,521	49,549	49,4925	0,002010046	0,000905709	2,842553923
(10;1)	35	43,0731	43,1066	43,14	43,1731	43,2061	43,1398	0,002765041	0,00121867	82,822787
(2;2)	63,27	63,3131	63,3125	63,3119	63,3113	63,3106	63,3119	9,53E-07	1,54918E-05	0,00219338
(3;3)	63,4	62,5712	62,5422	62,5136	62,4854	62,4576	62,514	0,002019165	0,000718334	0,98326154
(9;3)	63,83	59,0832	59,1289	59,174	59,2184	59,2621	59,1733	0,005000374	0,001195254	27,11083811
30% de dados retirados										
(1;1)	51	48,4151	48,4463	48,4773	48,5079	48,5381	48,4769	0,00236556	0,0010033	7,959655253
(3;1)	66,4	64,1331	64,1011	64,0695	64,0382	64,0072	64,0698	0,002477083	0,000776643	6,789649538
(7;1)	61,75	65,7575	65,726	65,6946	65,6633	65,6321	65,6947	0,002455675	0,000754533	19,45327968
(10;1)	35	42,524	42,5581	42,592	42,6257	42,6593	42,5918	0,00285926	0,001255508	72,04752315
(2;2)	63,27	63,3661	63,363	63,3599	63,3568	63,3537	63,3599	2,41E-05	7,73601E-05	0,010126537
(5;2)	62,23	61,0643	61,0599	61,0556	61,0513	61,0471	61,0557	4,62E-05	0,000111359	1,72394799
(3;3)	63,4	63,1499	63,1364	63,123	63,1099	63,097	63,1233	0,000436687	0,000331406	0,096182745
(7;3)	46,33	65,5462	65,5622	65,578	65,5937	65,6091	65,5778	0,000618023	0,000379273	463,0997994
(8;3)	68,67	63,9069	63,8908	63,875	63,8594	63,844	63,8752	0,000616743	0,00038914	28,7380119
(9;3)	63,83	62,9754	63,0014	63,0271	63,0523	63,0772	63,0267	0,00162047	0,000638485	0,808273165

D8

Posição da observação excluída	VO	Imputações					\bar{Y}	Var	CV	acc
		1	2	3	4	5				
5% de dados retirados										
(10;1)	38,5	74,6245	74,6372	74,6499	74,6625	74,6753	74,6499	0,000402006	0,000268784	1633,517683
15% de dados retirados										
(1;1)	79	68,4493	68,4492	68,4492	68,4491	68,4489	68,4491	2,35E-08	2,21562E-06	139,1508084
(10;1)	38,5	69,7835	69,7859	69,7883	69,7907	69,7931	69,7883	1,45E-05	5,43749E-05	1223,697161
(2;2)	87,45	80,519	80,5167	80,5143	80,512	80,5097	80,5143	1,36E-05	4,57578E-05	60,12923812
(3;3)	77,8	79,438	79,4453	79,4526	79,4599	79,4672	79,4526	0,000132717	0,000145273	3,413991675
(9;3)	81,33	80,3275	80,3288	80,3301	80,3314	80,3328	80,3301	4,37E-06	2,59846E-05	1,249704375
30% de dados retirados										
(1;1)	79	65,5142	65,5152	65,5162	65,5171	65,518	65,5162	2,27E-06	2,29358E-05	227,2681029
(3;1)	78,8	75,9103	75,8993	75,8884	75,8775	75,8667	75,8885	0,000297361	0,000227103	10,59677407
(7;1)	81,33	64,2489	64,2592	64,2695	64,2797	64,2899	64,2695	0,000262746	0,000252168	363,8286471
(10;1)	38,5	69,2277	69,238	69,2484	69,2587	69,269	69,2483	0,000265929	0,000235864	1181,82732
(2;2)	87,45	80,878	80,8716	80,8652	80,8588	80,8524	80,8652	0,000102804	0,000125138	54,1995912
(5;2)	79,46	76,1892	76,1904	76,1916	76,1927	76,1939	76,1916	3,39E-06	2,42827E-05	13,35337847
(3;3)	77,8	83,2783	83,2827	83,287	83,2914	83,2957	83,287	4,72E-05	8,2582E-05	37,63428291
(7;3)	59,11	84,3539	84,354	84,3541	84,3543	84,3544	84,3541	5,33E-08	2,45826E-06	796,5832555
(8;3)	89,44	83,0982	83,1021	83,1061	83,11	83,1139	83,1061	3,86E-05	7,47712E-05	50,14853352
(9;3)	81,33	82,6735	82,6758	82,678	82,6802	82,6824	82,678	1,23E-05	4,24571E-05	2,271324923