

**DAIBOU ALASSANE**

**DESEMPENHO DA ANÁLISE DE REGRESSÃO LINEAR REALIZADA SOB O  
DELINEAMENTO EM BLOCOS CASUALIZADOS**

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

Orientador: José Ivo Ribeiro Júnior

**VIÇOSA – MINAS GERAIS  
2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

Alassane, Daibou, 1985-

A323d Desempenho da análise de regressão linear realizada sob o  
2022 delineamento em blocos casualizados / Daibou Alassane. –  
Viçosa, MG, 2022.

1 dissertação eletrônica (60 f.): il. (algumas color.).

Orientador: José Ivo Ribeiro Junior.

Dissertação (mestrado) - Universidade Federal de Viçosa,  
Departamento de Estatística, 2022.

Referências bibliográficas: f. 58-60.

DOI: <https://doi.org/10.47328/ufvbbt.2022.132>

Modo de acesso: World Wide Web.

1. Análise de regressão. 2. Análise de erros (Matemática).  
3. Modelos matemáticos. I. Junior, José Ivo Ribeiro, 1966-.  
II. Universidade Federal de Viçosa. Departamento de Estatística.  
Programa de Pós-Graduação em Estatística Aplicada e  
Biometria. III. Título.

CDD 22. ed. 519.536

Bibliotecário(a) responsável: Alice Regina Pinto CRB6 2523

DAIBOU ALASSANE

DESEMPENHO DA ANÁLISE DE REGRESSÃO LINEAR REALIZADA SOB O  
DELINEAMENTO EM BLOCOS CASUALIZADOS

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

APROVADA: 21 de fevereiro de 2022.

Assentimento:

---

Daibou Alassane  
Autor

---

José Ivo Ribeiro Júnior  
Orientador

A Deus, meus familiares e  
aos meus amigos e  
companheiros

## AGRADECIMENTOS

A Deus, por estar sempre me indicando o melhor caminho, e me dando força em todos os momentos da vida, muito obrigado Senhor!

Ao orientador professor José Ivo Ribeiro Júnior, pela amizade, paciência, confiança, pelos conhecimentos transmitidos e pela orientação na realização deste trabalho.

Aos meus pais, Lamatou Soule e Mamam Alassane, meu tio Djibril Assouma, minhas irmãs e irmãos, pelo apoio, amizade, união e compreensão em todos os momentos.

Aos professores, colegas e amigos do departamento de estatística, pelos ensinamentos, disponibilidade, amizade e carinho.

A Jéssica Paula da Cruz, pelo apoio irrestrito, pelo incentivo, carinho e compreensão em todos os momentos.

Aos professores que participaram da banca examinadora, Rodrigo Luiz Pereira Lara, Eduardo Campana Barbosa, Danilo Pereira Barbosa e Paulo César Emiliano, por terem aceitado o convite e por suas contribuições oportunas, que certamente enriqueceram o trabalho.

À Universidade Federal de Viçosa, pela oportunidade de realização deste curso.

À CAPES pelo apoio financeiro.

Ao secretário do curso de pós-graduação em Estatística Aplicada e Biometria, Junior José Pires pelo apoio, dedicação, atenção e amizade.

A todas as pessoas que, direta ou indiretamente, estiveram presentes na realização deste trabalho.

## RESUMO

ALASSANE, Daibou, M.Sc., Universidade Federal de Viçosa, fevereiro de 2022. **Desempenho da análise de regressão linear realizada sob o delineamento em blocos casualizados.** Orientador: José Ivo Ribeiro Júnior.

Nas ciências agrárias, muitos experimentos são conduzidos com um ou dois fatores quantitativos sob o delineamento em blocos casualizados (DBC) e com quatro repetições por tratamento, cujas respostas são analisadas por meio da análise de regressão linear. O objetivo deste trabalho foi de avaliar os efeitos dos números de tratamentos e de repetições sobre o desempenho do modelo de regressão linear com uma e duas variáveis independentes e com dados coletados de experimentos instalados sob o DBC. Inicialmente, foi estabelecida uma equação de regressão linear simples para o estudo de uma variável e outra equação de regressão linear múltipla para o estudo de duas variáveis independentes. Em seguida, foram realizadas simulações de acordo com a distribuição normal para os erros do modelo de regressão com média populacional igual a zero e desvios-padrão populacionais para fornecerem diferentes precisões proporcionadas pelos respectivos coeficientes de variação iguais a 10, 20 e 30%. Além disso, foram realizadas três simulações para cada desvio-padrão, separadamente. No total, foram gerados 75.000 conjuntos de dados para o estudo da análise de regressão linear simples e 15.000 para o estudo da análise de regressão linear múltipla. E por fim, para cada uma das medidas avaliadas para verificar o desempenho dos modelos de regressão em função dos diferentes números de tratamentos e de repetições, foi realizada uma análise de superfície de resposta. Para o ajuste de um modelo de regressão linear simples em um experimento instalado sob o DBC, concluiu-se que para um mesmo número de unidades experimentais, o melhor é planejar o menor número possível de níveis quantitativos. Se houver uma expectativa para o modelo linear, pode-se então, recomendar apenas dois níveis quantitativos. Caso contrário, recomendam-se três. Para o ajuste de um modelo de regressão linear múltipla com duas variáveis independentes em um experimento instalado sob o DBC, concluiu-se, do mesmo modo, que para um mesmo número de unidades experimentais, o melhor é planejar, também, o menor número possível de combinações entre os níveis quantitativos das duas variáveis independentes. Se houver uma expectativa para o modelo com apenas efeitos lineares, pode-se então, recomendar apenas dois níveis quantitativos por variável independente avaliados em um fatorial  $2 \times 2$ . Caso contrário, recomendam-se três níveis por variável avaliados em um fatorial

3 × 3. Em ambos os casos, todos os tratamentos avaliados com o maior número possível de repetições.

Palavras-chave: Tratamentos. Repetições. Precisão experimental.

## ABSTRACT

ALASSANE, Daibou, M.Sc., Universidade Federal de Viçosa, February, 2022. **Performance of linear regression analysis accomplish under a randomized block design.** Adviser: José Ivo Ribeiro Júnior.

In agricultural sciences, many experiments are conducted with one or two quantitative factors under a randomized block design (RBD) and with four replications per treatment, whose responses are analyzed using linear regression analysis. Thus, the objective of this work was to evaluate the effects of the number of treatments and repetitions on the performance of the linear regression model with one and two independent variables and with data collected from experiments installed under RBD. Initially, a simple linear regression equation was established for the study of one variable and another multiple linear regression equation for the study of two independent variables. Then, several simulations were performed according to the normal distribution for the errors of the regression model with population mean equal to zero and population standard deviation to provide different accuracies provided by the respective coefficients of variation equal to 10, 20 and 30%. In addition, three simulations were performed for each standard deviation, separately. In total, 75,000 datasets were generated for the study of single linear regression analysis and 15,000 for the study of multiple linear regression analysis. Finally, for each of the measures evaluated to verify the performance of the regression models as a function of the different numbers of treatments and repetitions, a response surface analysis was performed. For the adjustment of a simple linear regression model in an experiment installed under RBD, it was concluded that for the same number of experimental units, it is best to plan the smallest possible number of quantitative levels. If there is an expectation for the linear model, then one can recommend only two quantitative levels. Otherwise, three are recommended. For the adjustment of a multiple linear regression model with two independent variables in an experiment installed under RBD, it was concluded, in the same way, that for the same number of experimental units, it is best to plan, also, the smallest possible number of combinations between the quantitative levels of the two independent variables. If there is an expectation for the model with only linear effects, then one can recommend only two quantitative levels per independent variable evaluated in a  $2 \times 2$  factorial. Otherwise, three levels per variable evaluated in a  $3 \times 3$  factorial are recommended. In both cases, all treatments were evaluated with the greatest possible number of repetitions.

Keywords: Treatments. Repetitions. Experimental precision.

## SUMÁRIO

1. INTRODUÇÃO.....	10
2. OBJETIVOS.....	12
2.1. Geral .....	12
3. REVISÃO DE LITERATURA .....	13
3.1. Regressão linear.....	13
3.2. Regressão linear simples.....	14
3.2.1. Delineamento em blocos casualizados.....	14
3.2.2. Modelo de regressão .....	18
3.3. Regressão linear múltipla .....	21
3.3.1. Fatorial completo $t \times g$ .....	21
3.3.2. Delineamento em blocos casualizados.....	22
3.3.3. Modelo de regressão .....	25
3.3.4. Delineamento composto central.....	28
3.3.5. Fatorial $2^2 + 1$ .....	30
3.4. Pressuposições .....	31
4. MATERIAL E MÉTODOS.....	32
4.1. Regressão linear simples.....	32
4.1.1. Parâmetros da regressão.....	32
4.1.2. Simulação dos dados.....	32
4.1.3. Delineamento em blocos casualizados.....	33
4.1.4. Medidas avaliadas .....	38
4.2. Regressão linear múltipla .....	40
4.2.1. Parâmetros da regressão.....	40
4.2.2. Simulação dos dados.....	41
4.2.3. Delineamento em blocos casualizados.....	41
4.2.4. Medidas avaliadas .....	45
5. RESULTADOS E DISCUSSÃO .....	48
5.1. Regressão linear simples.....	48
5.2. Regressão linear múltipla .....	53
6. CONCLUSÕES.....	57
7. REFERÊNCIAS BIBLIOGRÁFICAS .....	58

## 1. INTRODUÇÃO

Para o ajuste do modelo de regressão linear simples, deve-se, *a priori*, determinar a variável dependente (Y) e o intervalo dos valores da variável independente (X) definido pelos limites inferior e superior, quando se considera apenas um fator de interesse. Em seguida, determinar os seus níveis quantitativos (tratamentos), o número de repetições por tratamento e o delineamento de casualização. De forma geral, o modelo linear é composto pela média de Y que é uma função linear de X e pelo erro aleatório. Para ajustar o modelo linear, os erros devem ser uma variável aleatória normal e independentemente distribuída com média zero e variação constante. Além disso, deve-se assumir que exista uma relação linear entre Y e X (HOFFMANN; VIEIRA, 1983; COCHRAN; COX, 1992; DRAPER; SMITH, 1998; BARROS NETO et al., 2003; BANZATTO; KRONKA, 2006; WERKEMA; AGUIAR, 2006; GOMES, 2009; MONTGOMERY, 2009; DEVORE, 2012; MONTGOMERY; RUNGER, 2012; RIBEIRO JÚNIOR, 2012; ZIMMERMANN, 2014; GUPTA; GUTTMAN, 2017).

Quando se quer avaliar o efeito de uma variável independente e quantitativa X sobre uma variável dependente e quantitativa Y, o correto é estimar uma função, que entre elas, pode ser linear, e visualizar a relação em um diagrama de dispersão, com as devidas pressuposições satisfeitas e inferências estatísticas realizadas. Já para k variáveis independentes  $X_1, X_2, \dots, X_k$ , ajusta-se uma equação que mostra a relação de Y em função delas.

De acordo com uma comparação entre as aplicações das análises de regressão, estatísticas descritivas, métodos não paramétricos, testes de comparações de médias e análises multivariadas, em 1.237 artigos publicados na revista *Acta Scientiarum Agronomy* entre os anos de 1998 e 2016, as análises de regressão foram utilizadas em mais de 31% deles. Possato et al. (2019), verificaram também que os experimentos instalados sob o delineamento inteiramente casualizado (DIC) ocorreram em 31% das vezes. Sob o delineamento em blocos casualizados (DBC), em 43%. E no que diz respeito ao número de repetições por tratamento, 24% deles foram realizados com até três e 63% com quatro a seis. Em todos eles, os experimentos foram planejados de forma balanceada, isto é, com o mesmo número de repetições por tratamento.

Já em outro levantamento observacional realizado em 95 artigos relacionados a pesquisas no Brasil com as culturas do feijão, do milho e da soja, referentes a fatores com níveis quantitativos e publicados entre os anos de 2010 e 2017, constatou-se que os experimentos foram conduzidos, principalmente, com um e dois fatores sob o DIC e o DBC com três e quatro repetições e, com no máximo, oito repetições por tratamento. Experimentos com um e dois fatores com níveis quantitativos ocorreram em 48% e 46% das vezes, respectivamente. Sob o

DIC e DBC, os experimentos instalados foram realizados em 17% e 83% das vezes, respectivamente. E com três e quatro repetições por tratamento, foram conduzidos 16% e 67% dos experimentos, respectivamente. Nesse levantamento, os fatores com níveis quantitativos se referiram, principalmente, às seguintes áreas das Ciências Agrárias: nutrição e fertilidade do solo, fisiologia vegetal, fitopatologia, plantas daninhas, tratamento de sementes, microbiologia, irrigação e forragicultura.

Conforme relatado por Mateus et al. (2001), os experimentos realizados em laboratórios de fitopatologia e de entomologia apresentam precisões altas com, muitas vezes, coeficientes de variação inferiores a 6%. Por outro lado, em condições de campo, as precisões experimentais diminuem, proporcionando, na maioria das vezes, coeficientes de variação acima de 10%.

Como se pode observar, a frequência com que os dados experimentais, com diferentes precisões, são analisados por meio da análise de regressão, entre elas, a análise de regressão linear, é bastante alta. Além disso, percebe-se que os experimentos são instalados, principalmente, sob o DBC com quatro repetições por tratamento. Porém, conforme relatam Montgomery (2009) e Montgomery e Runger (2012), os planejamentos dos experimentos podem ser realizados como outros delineamentos experimentais e com outros números de tratamentos e repetições, a fim de possibilitarem experimentos mais adequados às suas respectivas precisões inerentes aos interesses ou condições técnicas e, conseqüentemente, com menores relações custo/benefício.

## **2. OBJETIVOS**

### **2.1. Geral**

Avaliar os efeitos dos números de tratamentos e de repetições sobre o desempenho do modelo de regressão linear com uma e duas variáveis independentes e com dados coletados de experimentos instalados sob o DBC.

### **2.2. Específicos**

Avaliar o desempenho do modelo de regressão linear com uma e duas variáveis independentes e com dados simulados de experimentos instalados sob o DBC com diferentes precisões proporcionadas pelos respectivos coeficientes de variação;

Verificar, para um mesmo número de unidades experimentais, se é preferível experimentar mais níveis quantitativos com menos repetições ou menos níveis quantitativos com mais repetições; e

Verificar a possibilidade da economia do planejamento experimental sob o DBC proporcionada pela redução do número de tratamentos e/ou de blocos (repetições).

### 3. REVISÃO DE LITERATURA

#### 3.1. Regressão linear

O termo regressão tem uma história interessante, que remonta ao trabalho de Sir Francis Galton em 1885, a partir da investigação da relação entre as alturas dos pais e dos filhos. Ele descobriu, conforme esperado, que os pais mais altos tendem a ter filhos mais altos, e que os pais mais baixos tendem a ter filhos mais baixos. Contudo, descobriu, também, que os filhos dos pais muito altos eram mais baixos do que os seus pais, e que os filhos dos pais muito baixos eram mais altos do que os seus pais. Galton chamou esse fato de “regressão” em direção à média.

De acordo com Gea et al. (2013), frequentemente são realizadas associações entre duas ou mais variáveis, com o objetivo de indagar se elas são inter-relacionadas. A correlação e a regressão são conceitos estatísticos fundamentais que se estendem à ideia de dependência funcional, exigindo o conhecimento de conceitos sobre covariância, distribuição de probabilidades, centralização e dispersão.

Segundo Krajewski et al. (2009), o modelo de regressão linear é um dos modelos estatísticos mais conhecidos e utilizados, que consiste de uma variável chamada dependente relacionada a uma ou mais variáveis independentes por meio de uma equação linear nos parâmetros.

A regressão é dita ser linear porque considera que a relação da variável dependente às variáveis independentes é uma função linear dos parâmetros. Portanto, o objetivo da análise de regressão linear é verificar a existência de uma relação funcional e linear entre uma variável dependente  $Y$  com uma (simples) ou com duas ou mais variáveis independentes  $X_s$  (múltipla). A diferença entre uma regressão linear simples e uma regressão linear múltipla está na quantidade de variáveis independentes avaliadas.

Para Maroco (2010), a regressão linear é definida como um conjunto de métodos estatísticos que são utilizados para modelar relações entre variáveis. E também, para prever a média de uma variável dependente condicionada a um conjunto de variáveis independentes.

Desse modo, para estimar o valor esperado de  $Y$  em função de  $X_1$  (regressão linear simples) ou em função de  $X_1, X_2, \dots, X_k$  (regressão linear múltipla com  $k$  variáveis independentes), utiliza-se de uma equação que determina a relação entre ambos os tipos de variáveis. Consequentemente, de uma equação que determina a relação de uma variável

dependente com uma variável independente ou de uma variável dependente com  $k$  variáveis independentes.

## 3.2. Regressão linear simples

### 3.2.1. Delineamento em blocos casualizados

Segundo Gomes (2009), o delineamento em blocos casualizados (DBC) constitui o delineamento de casualização mais importante nos experimentos realizados nas áreas de Ciências Agrárias. Por exemplo, se é de interesse estudar a adubação dos canaviais de uma usina de açúcar, escolhe-se para estar em cada bloco, uma parte uniforme do terreno. Além disso, pode-se espalhar os blocos por toda a propriedade, obtendo, assim, conclusões válidas para toda a área cultivada e não somente para uma determinada parte uniforme.

O DBC é o delineamento que incorpora, além do efeito do fator de interesse em um experimento com um fator (variável independente  $X_1$ ), o efeito de um fator perturbador para a constituição dos blocos. Nesse caso, a casualização dos tratamentos (níveis quantitativos de  $X_1$ ) é feita dentro de cada bloco, separadamente. Portanto, no DBC faz-se o uso dos três princípios básicos da experimentação: repetição, casualização e controle local. Por outro lado, a utilização do controle local sempre conduz a uma redução no número de graus de liberdade do resíduo. E tal redução é a principal desvantagem de usar o DBC e não o DIC.

Para um experimento balanceado de  $t$  tratamentos e  $r$  repetições, cujo fator de interesse é a própria variável independente  $X_1$ , o modelo estatístico sob o DBC é definido por:

$$y_{ij} = \mu + \tau_i + \omega_j + \varepsilon_{ij}, \text{ em que:}$$

$y_{ij}$ : valor observado da variável dependente  $Y$  no nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e no bloco  $b_j$  ( $j = 1, 2, \dots, r$ );

$\mu$ : média geral populacional da variável dependente  $Y$ ;

$\tau_i = \mu_i - \mu$ : efeito do nível quantitativo  $x_{1i}$ ;

$\omega_j = \mu_j - \mu$ : efeito do bloco  $b_j$ ;

$\varepsilon_{ij} = y_{ij} - \mu_i - \mu_j + \mu$ : erro experimental associado ao valor observado  $y_{ij}$ ;

$\mu_i$ : média populacional da variável dependente  $Y$  no nível quantitativo  $x_{1i}$ ; e

$\mu_j$ : média populacional da variável dependente  $Y$  no bloco  $b_j$ .

Quando os níveis de  $X_1$  são quantitativos, não se aplica o teste F da análise de variância (ANOVA) para a fonte de variação devida aos tratamentos. Portanto, não são testadas as hipóteses relacionadas às suas médias, como segue:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_t; e$$

$$H_1: \text{pelo menos uma média } \mu_i \text{ do nível quantitativo } x_{1i} \text{ difere das demais (} i = 1, 2, \dots, t \text{)}.$$

Nesse caso, os efeitos dos níveis quantitativos da variável independente  $X_1$  são testados pelo teste F da ANOVA sobre o parâmetro do modelo de regressão linear simples que se refere ao coeficiente da regressão ( $\beta_1$ ), conforme as seguintes hipóteses:

$$H_{0_1}: \beta_1 = 0; e$$

$$H_{1_1}: \beta_1 \neq 0.$$

Porém, quando os níveis quantitativos da variável independente  $X_1$  são repetidos duas ou mais vezes em um experimento instalado sob o DBC com um total de  $tr$  unidades experimentais, isto é, quando há para cada nível quantitativo  $x_{1i}$  mais de um valor observado da variável dependente  $Y$ , a variância dos erros experimentais comum em todos os níveis quantitativos de  $X_1$  ( $\sigma_\varepsilon^2$ ) poderá ser obtida a partir de dois estimadores, como segue:

QMRes: quadrado médio do resíduo; e

QMResReg: quadrado médio do resíduo da regressão (ResReg).

O primeiro estimador, também denominado de quadrado médio do resíduo puro, é obtido pela ANOVA de acordo com o modelo estatístico baseado no DBC. Ele representa a variância dos erros experimentais das diferentes repetições (blocos) de um mesmo nível quantitativo de  $X_1$  e comum a todos eles (QMRes).

Já o QMResReg obtido pela ANOVA da regressão é composto, no DBC, pelo QMRes, pela falta de ajuste das médias dos níveis quantitativos de  $X_1$  à regressão e pelos efeitos de blocos. Portanto, se o parâmetro  $\beta_1$  do modelo de regressão for testado pelo QMResReg, este não deverá conter falta de ajuste e nem efeitos de blocos significativos. Caso contrário, ele estará incluindo um efeito não aleatório devido ao uso de um modelo inadequado com a presença da falta de ajuste e/ou dos efeitos de blocos significativos.

As hipóteses testadas pelo teste F da ANOVA sobre a falta de ajuste das médias dos níveis quantitativos de  $X_1$  à regressão são obtidas por:

$$H_{0_2}: \text{não há falta de ajuste (modelo de regressão adequado); e}$$

$$H_{1_2}: \text{há falta de ajuste (modelo de regressão não adequado).}$$

Já as hipóteses testadas, também, pelo teste F da ANOVA sobre os efeitos de blocos são dadas por:

$H_{0_3}$ :  $\mu_1 = \mu_2 = \dots = \mu_r$ ; e

$H_{1_3}$ : pelo menos uma média  $\mu_j$  do bloco  $b_j$  difere das demais ( $j = 1, 2, \dots, r$ ).

No entanto, como os efeitos de blocos tendem a ser significativos, ou seja, tende a ocorrer a decisão de rejeitar  $H_{0_3}$ , pois foi em função deles a opção de utilizar o DBC, é recomendado retirar do QMResReg, os efeitos de blocos. Além disso, ajustar adequadamente um modelo de regressão para que não ocorra a falta de ajuste, ou seja, para que nesse último caso não se rejeite  $H_{0_2}$ . Nessas condições, o QMResReg e o QMRes serão, igualmente, bons estimadores da variância dos erros experimentais das diferentes repetições (blocos) de um mesmo nível quantitativo de  $X_1$  e comum a todos eles. Consequentemente, o parâmetro  $\beta_1$  ( $H_{0_1}$ ) do modelo de regressão linear simples poderá ser testado pelo QMResReg (Tabela 3.1).

Tabela 3.1 - Análise de variância da regressão com o teste para a falta de ajuste sob o DBC de um experimento balanceado

FV	GL	SQ	QM	F
Bloco	$r - 1$	SQBI	$SQBI/(r - 1)$	$QMBI/QMResReg$
Regressão	1	SQReg	$SQReg/1$	$QMReg/QMResReg$
ResReg	$r(t - 1) - 1$	SQResReg	$SQResReg/[r(t - 1) - 1]$	
Falta de Ajuste	$t - 2$	SQFA	$SQFA/(t - 2)$	$QMFA/QMRes$
Resíduo	$(t - 1)(r - 1)$	SQRes	$SQRes/[(t - 1)(r - 1)]$	
Total	$tr - 1$	SQTotal		

Em termos matriciais e de somatórios, as fórmulas práticas para obter as somas de quadrados da ANOVA são dadas por:

$$SQBI = \frac{1}{t} \sum_{j=1}^r B_j^2 - C;$$

$$SQReg = \hat{\beta}' \mathbf{X}' \mathbf{y} - C;$$

$$SQResReg = SQTotal - SQBI - SQReg = SQFA + SQRes;$$

$$SQFA = SQTrat - SQReg;$$

$$SQRes = SQTotal - SQBI - SQTrat; e$$

$$SQTotal = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - C, \text{ em que:}$$

$$SQTrat = \frac{1}{r} \sum_{i=1}^t X_{1_i}^2 - C;$$

$$B_j = \sum_{i=1}^t y_{ij}; \text{ total amostral da variável dependente } Y \text{ do bloco } b_j \text{ (} j = 1, 2, \dots, r \text{);}$$

$X_{1i} = \sum_{j=1}^r y_{ij}$ : total amostral da variável dependente  $Y$  do nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ );

$$C = \frac{(\sum_{i=1}^t \sum_{j=1}^r y_{ij})^2}{tr};$$

$\mathbf{y}$ : vetor  $tr \times 1$  dos valores observados da variável dependente  $Y$ ;

$\mathbf{X}$ : matriz de incidência  $tr \times 2$ , sendo a primeira coluna composta pela constante 1, e a segunda, pelos níveis quantitativos de  $X_1$ ; e

$\hat{\boldsymbol{\beta}}$ : vetor  $2 \times 1$  das estimativas dos parâmetros  $\beta_0$  e  $\beta_1$  do modelo de regressão linear simples.

Se  $f_{\text{cal}_{\text{BI}}} \geq f_{\text{tab}_{\text{BI}}}$  (valor  $p \leq \alpha$ ), rejeita-se  $H_{03}$ , sendo o  $f_{\text{tab}_{\text{BI}}}$  dado por:

$f_{\alpha, r-1, r(t-1)-1}$ : valor da distribuição  $F$  com  $r - 1$  graus de liberdades (gl) de blocos e  $r(t - 1) - 1$  gl do resíduo da regressão que deixa uma probabilidade  $\alpha$  na extremidade da cauda à direita.

Do mesmo modo, se  $f_{\text{cal}_{\text{Reg}}} \geq f_{\text{tab}_{\text{Reg}}}$  (valor  $p \leq \alpha$ ), rejeita-se  $H_{01}$ , sendo o  $f_{\text{tab}_{\text{Reg}}}$  dado por:

$f_{\alpha, 1, r(t-1)-1}$ : valor da distribuição  $F$  com 1 gl da regressão e  $r(t - 1) - 1$  gl do resíduo da regressão que deixa uma probabilidade  $\alpha$  na extremidade da cauda à direita.

Como já mencionado, é importante ajustar um modelo de regressão adequado para que a hipótese  $H_{02}$  sobre a falta de ajuste não seja rejeitada. Assim, se  $f_{\text{cal}_{\text{FA}}} < f_{\text{tab}_{\text{FA}}}$  (valor  $p > \alpha$ ), não se rejeita  $H_{02}$ , sendo o  $f_{\text{tab}_{\text{FA}}}$  dado por:

$f_{\alpha, t-2, (t-1)(r-1)}$ : valor da distribuição  $F$  com  $t - 2$  gl da falta de ajuste e  $(t - 1)(r - 1)$  gl do resíduo (puro) que deixa uma probabilidade  $\alpha$  na extremidade da cauda à direita.

Quando se tem 1 gl para a regressão, os testes  $F$  e  $t$  de Student são equivalentes. Portanto, ao invés do primeiro, pode-se aplicar o teste  $t$  de Student para testar  $H_{01}: \beta_1 = 0$  vs.  $H_{11}: \beta_1 \neq 0$ , cujo cálculo é obtido por:

$$t_{\text{cal}_{b1}} = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}.$$

Se  $|t_{\text{cal}_{b1}}| \geq t_{\text{tab}_{b1}}$  (valor  $p \leq \alpha$ ), rejeita-se  $H_{01}$ , sendo o  $t_{\text{tab}_{b1}}$  dado por:

$t_{\alpha/2, r(t-1)-1}$ : valor da distribuição  $t$  de Student com  $r(t - 1) - 1$  gl do resíduo da regressão que deixa uma probabilidade  $\alpha/2$  na extremidade da cauda à direita.

Portanto, nessa última situação, tem-se:

$$t_{\text{cal}_{b1}}^2 = f_{\text{cal}_{\text{Reg}}}; \text{ e}$$

$$t_{\text{tab}_{b1}}^2 = f_{\text{tab}_{\text{Reg}}}.$$

De acordo com Shieh e Jan (2004), pode-se medir a eficiência relativa, isto é, o quanto se perderia ou se ganharia se fosse utilizado o DIC no lugar do DBC. Isso significa que o único

delineamento utilizado para a instalação do referido experimento foi o DBC. A medida de eficiência relativa (ER) mais amplamente utilizada considera apenas as variâncias residuais estimadas pelas ANOVA baseadas nos modelos estatísticos dos dois delineamentos (DIC e DBC), quando os valores observados foram obtidos de um experimento instalado sob o DBC, como segue:

$$ER = \frac{QMRes_{DIC}}{QMRes} = \frac{(r-1)QMBI + r(t-1)QMRes}{(tr-1)QMRes}, \text{ em que:}$$

$QMRes_{DIC}$ : quadrado médio do resíduo do DIC;

$QMRes$ : quadrado médio do resíduo do DBC;

$QMBI$ : quadrado médio de blocos do DBC;

$r$ : número de blocos do DBC; e

$t$ : número de tratamentos do DBC.

A eficiência relativa é um número positivo que pode ser interpretado, de forma proporcional, como o número de unidades experimentais que seria utilizado no DIC que acarretaria a mesma eficiência do DBC. Isso implica que serão necessárias  $r \times ER$  repetições por tratamento no DIC, para que ele proporcione o mesmo grau de precisão associado ao  $QMRes$  estimado pelo DBC.

### 3.2.2. Modelo de regressão

O modelo de regressão linear simples de primeiro grau para estudar  $Y$  em função de  $X_1$  em um experimento instalado sob o DBC com  $tr$  unidades experimentais compostas por  $t$  níveis quantitativos de  $X_1$  e  $r$  blocos, e sem a consideração do modelo estatístico do DBC, é dado por:

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \varepsilon'_{ij}, \text{ em que:}$$

$y_{ij}$ : valor observado da variável dependente  $Y$  no nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e no bloco  $b_j$  ( $j = 1, 2, \dots, r$ );

$\beta_0$ : constante da regressão;

$\beta_1$ : coeficiente da regressão;

$\varepsilon'_{ij} = y_{ij} - \mu_i = \varepsilon_{ij} + \phi_i$ : erro da regressão associado ao valor observado  $y_{ij}$ ;

$\mu_i = \beta_0 + \beta_1 x_{1i}$ : média populacional da variável dependente  $Y$  no nível quantitativo  $x_{1i}$ ; e

$\phi_i$ : falta de ajuste no nível quantitativo  $x_{1i}$ .

O modelo de regressão linear simples foi definido conforme a expressão apresentada por Bastos et al. (2015).

Assim, para os resultados do experimento com  $tr$  pares de valores  $(x_{1i}$  e  $y_{ij})$ , conclui-se que os de  $Y$  dependem dos de  $X_1$  e dos erros da regressão  $(\varepsilon'_{ij})$ , como segue:

$$y_{11} = \beta_0 + \beta_1 x_{11} + \varepsilon'_{11};$$

$$y_{12} = \beta_0 + \beta_1 x_{12} + \varepsilon'_{12};$$

...

$$y_{tr} = \beta_0 + \beta_1 x_{1t} + \varepsilon'_{tr}.$$

Em termos matriciais, o modelo linear  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$  é expresso por:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \dots \\ y_{1r} \\ y_{21} \\ y_{22} \\ \dots \\ y_{2r} \\ \dots \\ y_{t1} \\ y_{t2} \\ \dots \\ y_{tr} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ \dots & \dots \\ 1 & x_{1r} \\ 1 & x_{12} \\ 1 & x_{12} \\ \dots & \dots \\ 1 & x_{12} \\ \dots & \dots \\ 1 & x_{1t} \\ 1 & x_{1t} \\ \dots & \dots \\ 1 & x_{1t} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon'_{11} \\ \varepsilon'_{12} \\ \dots \\ \varepsilon'_{1r} \\ \varepsilon'_{21} \\ \varepsilon'_{22} \\ \dots \\ \varepsilon'_{2r} \\ \dots \\ \varepsilon'_{t1} \\ \varepsilon'_{t2} \\ \dots \\ \varepsilon'_{tr} \end{bmatrix}, \text{ em que:}$$

$\mathbf{y}$ : vetor  $tr \times 1$  dos valores observados da variável dependente  $Y$ ;

$\mathbf{X}$ : matriz de incidência  $tr \times 2$ , sendo a primeira coluna composta pela constante 1, e a segunda, pelos níveis quantitativos de  $X_1$ ;

$\boldsymbol{\beta}$ : vetor  $2 \times 1$  dos parâmetros  $\beta_0$  e  $\beta_1$ ; e

$\boldsymbol{\varepsilon}'$ : vetor  $tr \times 1$  dos erros da regressão.

Entre 1777 e 1855, conforme citado por Maroco (2010), Karl Gauss propôs estimar os parâmetros  $\beta_0$  e  $\beta_1$  visando minimizar a soma dos quadrados dos erros da regressão. Este processo ficou conhecido como o método de mínimos quadrados.

O primeiro passo na análise de regressão consiste em obter as estimativas dos parâmetros  $\beta_0$  e  $\beta_1$ . Para isso, o método de mínimos quadrados resulta nas estimativas que minimizam a soma de quadrados dos erros da regressão. Em termos matriciais, deve-se minimizar:

$$L = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

Portanto, o vetor  $\hat{\boldsymbol{\beta}}$  dos estimadores de mínimos quadrados tem de satisfazer à seguinte condição:

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}.$$

Matricialmente, o sistema de equações normais  $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$  é dado por:

$$\begin{bmatrix} \text{tr} & r \sum_{i=1}^t x_{1i} \\ r \sum_{i=1}^t x_{1i} & r \sum_{i=1}^t x_{1i}^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^t \sum_{j=1}^r y_{ij} \\ \sum_{i=1}^t \sum_{j=1}^r x_{1i} y_{ij} \end{bmatrix}.$$

Assim, após a resolução, o vetor  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , apresenta as estimativas dos parâmetros  $\beta_0$  e  $\beta_1$ .

A equação de regressão ajustada é apropriada para estimar as médias da variável dependente  $Y$  em função de quaisquer níveis quantitativos da variável independente  $X_1$  que estejam dentro do intervalo estudado. Para o modelo de regressão linear de primeiro grau, tem-se:

$$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i}, \text{ em que:}$$

$\hat{y}_{ij}$ : valor ajustado da variável dependente  $Y$  no nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e no bloco  $b_j$  ( $j = 1, 2, \dots, r$ ).

Em termos de média, os estimadores de mínimos quadrados  $\hat{\beta}_0$  e  $\hat{\beta}_1$  são não tendenciosos, uma vez que:

$$E(\hat{\beta}_0) = \beta_0; \text{ e}$$

$$E(\hat{\beta}_1) = \beta_1.$$

Já as estimativas das variâncias dos estimadores de mínimos quadrados  $\hat{\beta}_0$  e  $\hat{\beta}_1$  são obtidas na diagonal da seguinte matriz:

$$\hat{V}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} s^2(\hat{\beta}_0) & s(\hat{\beta}_0, \hat{\beta}_1) \\ s(\hat{\beta}_0, \hat{\beta}_1) & s^2(\hat{\beta}_1) \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \text{QMResReg}, \text{ em que:}$$

QMResReg: quadrado médio do resíduo da regressão (Tabela 3.1).

As estimativas dos desvios-padrão dos estimadores  $\hat{\beta}_0$  e  $\hat{\beta}_1$ , fornecidas por  $s(\hat{\beta}_0)$  e  $s(\hat{\beta}_1)$ , são obtidas extraindo-se a raiz quadrada das respectivas estimativas das variâncias.

### 3.3. Regressão linear múltipla

#### 3.3.1. Fatorial completo $t \times g$

Um experimento é somente um teste ou uma série de testes. Segundo Gomes (2009), experimentos fatoriais completos são aqueles que incluem todas as combinações entre os níveis de todos os fatores estudados. Portanto, em um fatorial completo  $t \times g$ , isto é, com  $t$  níveis do primeiro fator (variável independente  $X_1$ ) e  $g$  níveis do segundo fator (variável independente  $X_2$ ), são avaliadas  $tg$  combinações entre eles ou tratamentos.

Nos experimentos fatoriais completos o número de tratamentos é geralmente elevado, pois envolve todas as combinações possíveis entre os níveis dos fatores investigados. O fatorial é um delineamento de tratamentos e não um delineamento de casualização, que representa a maneira pela qual eles são casualizados às unidades experimentais. Portanto, os experimentos fatoriais são montados segundo algum tipo de delineamento de casualização e, quando todos os fatores (variáveis independentes) possuem níveis quantitativos, eles podem fazer parte da metodologia de superfície de resposta.

A superfície de resposta é um método de estimação que se utiliza de planos adequados para a coleta de dados, cujo modelo estatístico é representado, de forma empírica, por um polinômio de primeira ou de segunda ordem. Desse modo, tem-se:  $Y = f(X_1, X_2, \dots, X_k) + \varepsilon$ , em que  $f$  é a função que modela a relação de dependência da variável dependente  $Y$  pelas variáveis independentes  $X_1, X_2, \dots, X_k$ .

Nos experimentos fatoriais, podem ser estudados os efeitos principais e das interações. O efeito principal é o efeito de cada variável independente que ocorre independentemente dos efeitos das outras variáveis independentes. E o efeito de interação é o efeito simultâneo das variáveis independentes que interagem sobre a variável dependente em estudo.

De acordo com Bonilla (1995), o objetivo mais importante dos experimentos fatoriais é a possibilidade de avaliar as interações. Uma interação dupla pode ser definida como a falha ou o fracasso dos níveis de uma variável independente em manter o mesmo grau e a magnitude de efeitos, quando combinados com todos os níveis de outra variável independente. No mesmo sentido, Gomes (2009) relata que a presença da interação indica que o comportamento de uma variável independente depende dos níveis das outras variáveis independentes. Do mesmo modo, segundo Mason et al. (2003) e Montgomery (2009), o efeito da interação ocorre quando o efeito

de uma variável independente sobre a variável dependente depende dos níveis das outras variáveis independentes.

Segundo Lima e Abreu (2000), os experimentos fatoriais permitem um estudo mais aprimorado das variáveis independentes, já que estando cada uma delas sob condições diferentes das outras, a análise é mais completa e as conclusões são mais abrangentes. Além disso, mostram como se devem ensaiar todas as combinações possíveis, possibilitam maior número de graus de liberdade do resíduo e fornecem uma estimativa da interação entre as variáveis independentes avaliadas.

No entanto, de acordo com Gomes (2009), a principal desvantagem dos experimentos fatoriais é que o número de tratamentos aumenta rapidamente. Portanto, um experimento fatorial completo instalado sob o DBC, com um grande número de tratamentos, pode ter a sua eficiência reduzida.

### 3.3.2. Delineamento em blocos casualizados

Para um experimento fatorial  $t \times g$  com dois fatores de interesse (variáveis independentes  $X_1$  e  $X_2$ ) instalado sob o DBC, além dos dois efeitos principais e das interações entre eles, há novamente a presença de um fator perturbador para a constituição dos blocos. Do mesmo modo, a casualização dos tratamentos (combinações entre os níveis de  $X_1$  e  $X_2$ ) é feita dentro de cada bloco, separadamente.

Desse modo, para um experimento fatorial balanceado, cujos fatores de interesse são as próprias variáveis independentes  $X_1$  e  $X_2$ , o modelo estatístico sob o DBC é definido por:

$$y_{ijk} = \mu + \tau_i + \gamma_j + \tau\gamma_{ij} + \omega_k + \varepsilon_{ijk}, \text{ em que:}$$

$y_{ijk}$ : valor observado da variável dependente  $Y$  na combinação entre os níveis quantitativos  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e  $x_{2j}$  ( $j = 1, 2, \dots, g$ ) e no bloco  $b_k$  ( $k = 1, 2, \dots, r$ );

$\mu$ : média geral populacional da variável dependente  $Y$ ;

$\tau_i = \mu_i - \mu$ : efeito do nível quantitativo  $x_{1i}$ ;

$\gamma_j = \mu_j - \mu$ : efeito do nível quantitativo  $x_{2j}$ ;

$\omega_k = \mu_k - \mu$ : efeito do bloco  $b_k$ ;

$\varepsilon_{ijk} = y_{ijk} - \mu_{ij} - \mu_k + \mu$ : erro experimental associado ao valor observado  $y_{ijk}$ ;

$\mu_i$ : média populacional da variável dependente  $Y$  no nível quantitativo  $x_{1i}$ ;

$\mu_j$ : média populacional da variável dependente  $Y$  no nível quantitativo  $x_{2j}$ ;

$\mu_{ij}$ : média populacional da variável dependente Y na combinação entre os níveis quantitativos  $x_{1i}$  e  $x_{2j}$ ; e

$\mu_k$ : média populacional da variável dependente Y no bloco  $b_k$ .

Do mesmo modo, quando as combinações entre os níveis de  $X_1$  e  $X_2$  são quantitativas, não se aplica o teste F da ANOVA para as fontes de variação devidas aos efeitos principais de  $X_1$  e  $X_2$  e nem aos efeitos da interação  $X_1 \times X_2$ . Portanto, não são testados os seguintes pares de hipóteses em relação aos três tipos de efeitos anteriores, respectivamente:

$H_{0_1}$ :  $\mu_1 = \mu_2 = \dots = \mu_t$  vs.  $H_{1_1}$ : pelo menos uma média  $\mu_i$  do nível quantitativo  $x_{1i}$  difere das demais ( $i = 1, 2, \dots, t$ );

$H_{0_2}$ :  $\mu_1 = \mu_2 = \dots = \mu_g$  vs.  $H_{1_2}$ : pelo menos uma média  $\mu_j$  do nível quantitativo  $x_{2j}$  difere das demais ( $j = 1, 2, \dots, g$ ); e

$H_{0_3}$ :  $\tau\gamma_{11} = \tau\gamma_{12} = \dots = \tau\gamma_{tg} = 0$  vs.  $H_{1_3}$ : pelo menos um efeito  $\tau\gamma_{ij}$  difere de zero, em que:  
 $\tau\gamma_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$ : efeito da interação entre os níveis quantitativos  $x_{1i}$  e  $x_{2j}$ .

Nesse caso, os efeitos principais dos níveis quantitativos das variáveis independentes  $X_1$  e  $X_2$ , dada a ausência da interação  $X_1 \times X_2$ , são testados pelo teste F da ANOVA sobre os parâmetros do modelo de regressão linear múltipla que se referem aos coeficientes da regressão ( $\beta_1$  e  $\beta_2$ ), conforme as seguintes hipóteses:

$H_{0_{12}}$ :  $\beta_1 = \beta_2 = 0$ ; e

$H_{1_{12}}$ :  $\beta_w \neq 0$ , para pelo menos um  $w$  ( $w = 1$  e  $2$ ).

As hipóteses testadas pelo teste F da ANOVA sobre a falta de ajuste das médias das combinações entre os níveis quantitativos de  $X_1$  e  $X_2$  à regressão são obtidas por:

$H_{0_3}$ : não há falta de ajuste (modelo de regressão adequado); e

$H_{1_3}$ : há falta de ajuste (modelo de regressão não adequado).

Já as hipóteses testadas, também, pelo teste F da ANOVA sobre os efeitos de blocos são dadas por:

$H_{0_4}$ :  $\mu_1 = \mu_2 = \dots = \mu_r$ ; e

$H_{1_4}$ : pelo menos uma média  $\mu_k$  do bloco  $b_k$  difere das demais ( $k = 1, 2, \dots, r$ ).

Do mesmo modo, como os efeitos de blocos tendem a ser significativos e como o objetivo é de ajustar adequadamente um modelo de regressão linear múltipla sem interação, o QMResReg e o QMRes serão, igualmente, bons estimadores, quando for retirado do primeiro, os efeitos de blocos. Nesse caso, os parâmetros  $\beta_1$  e  $\beta_2$  do modelo de regressão linear múltiplo poderão ser testados pelo QMResReg (Tabela 3.2).

Tabela 3.2 - Análise de variância da regressão com o teste para a falta de ajuste sob o DBC de um experimento fatorial balanceado

FV	GL	SQ	QM	F
Bloco	$r - 1$	SQBI	$SQBI/(r - 1)$	$QMBI/QMResReg$
Regressão	2	SQReg	$SQReg/2$	$QMReg/QMResReg$
ResReg	$r(tg - 1) - 2$	SQResReg	$SQResReg/[r(tg - 1) - 2]$	
Falta de Ajuste	$tg - 3$	SQFA	$SQFA/(tg - 3)$	$QMFA/QMRes$
Resíduo	$(tg - 1)(r - 1)$	SQRes	$SQRes/[(tg - 1)(r - 1)]$	
Total	$tgr - 1$	SQTotal		

Em termos matriciais e de somatórios, as fórmulas práticas para obter as somas de quadrados da ANOVA são dadas por:

$$SQBI = \frac{1}{tg} \sum_{k=1}^r B_k^2 - C;$$

$$SQReg = \hat{\beta}' \mathbf{X}' \mathbf{y} - C;$$

$$SQResReg = SQTotal - SQBI - SQReg = SQFA + SQRes;$$

$$SQFA = SQTrat - SQReg;$$

$$SQRes = SQTotal - SQBI - SQTrat; e$$

$$SQTotal = \sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r y_{ijk}^2 - C, \text{ em que:}$$

$$SQTrat = \frac{1}{r} \sum_{i=1}^t \sum_{j=1}^g (X_1 X_2)_{ij}^2 - C;$$

$B_j = \sum_{k=1}^r y_{ijk}$ : total amostral da variável dependente Y do bloco  $b_j$  ( $j = 1, 2, \dots, r$ );

$X_{1i} = \sum_{j=1}^g \sum_{k=1}^r y_{ijk}$ : total amostral da variável dependente Y do nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ );

$X_{2j} = \sum_{i=1}^t \sum_{k=1}^r y_{ijk}$ : total amostral da variável dependente Y do nível quantitativo  $x_{2j}$  ( $j = 1, 2, \dots, g$ );

$(X_1 X_2)_{ij}$ : total amostral da variável dependente Y da combinação entre os níveis quantitativos  $x_{1i}$  e  $x_{2j}$ ;

$$C = \frac{(\sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r y_{ijk})^2}{tgr};$$

$\mathbf{y}$ : vetor  $tgr \times 1$  dos valores observados da variável dependente Y;

$\mathbf{X}$ : matriz de incidência  $tgr \times 3$ , sendo a primeira coluna composta pela constante 1, a segunda, pelos níveis quantitativos de  $X_1$ , e a terceira, pelos níveis quantitativos de  $X_2$ ;

$\hat{\beta}$ : vetor  $3 \times 1$  das estimativas dos parâmetros  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  do modelo de regressão linear múltiplo.

Se  $f_{\text{cal}_{B1}} \geq f_{\text{tab}_{B1}}$  (valor  $p \leq \alpha$ ), rejeita-se  $H_{04}$ , sendo o  $f_{\text{tab}_{B1}}$  dado por:

$f_{\alpha, r-1, r(tg-1)-2}$ : valor da distribuição F com  $r - 1$  gl de blocos e  $r(tg - 1) - 2$  gl do resíduo da regressão que deixa uma probabilidade  $\alpha$  na extremidade da cauda à direita.

Do mesmo modo, se  $f_{\text{cal}_{Reg}} \geq f_{\text{tab}_{Reg}}$  (valor  $p \leq \alpha$ ), rejeita-se  $H_{012}$ , sendo o  $f_{\text{tab}_{Reg}}$  dado por:

$f_{\alpha, 2, r(tg-1)-2}$ : valor da distribuição F com 1 gl da regressão e  $r(tg - 1) - 2$  gl do resíduo da regressão que deixa uma probabilidade  $\alpha$  na extremidade da cauda à direita.

Como já mencionado, é importante ajustar um modelo de regressão adequado para que a hipótese  $H_{03}$  sobre a falta de ajuste não seja rejeitada. Assim, se  $f_{\text{cal}_{FA}} < f_{\text{tab}_{FA}}$  (valor  $p > \alpha$ ), não se rejeita  $H_{03}$ , sendo o  $f_{\text{tab}_{FA}}$  dado por:

$f_{\alpha, tg-3, (tg-1)(r-1)}$ : valor da distribuição F com  $tg - 3$  gl da falta de ajuste e  $(tg - 1)(r - 1)$  gl do resíduo (puro) que deixa uma probabilidade  $\alpha$  na extremidade da cauda à direita.

No entanto, quando se rejeita  $H_{012}$ , não se pode afirmar se é um ( $\beta_1$  ou  $\beta_2$ ) ou se são os dois coeficientes ( $\beta_1$  e  $\beta_2$ ) que diferem de zero, pode-se aplicar o teste t de Student para testar  $H_{01}$ :  $\beta_1 = 0$  vs.  $H_{11}$ :  $\beta_1 \neq 0$  e  $H_{02}$ :  $\beta_2 = 0$  vs.  $H_{12}$ :  $\beta_2 \neq 0$ , cujos cálculos são obtidos, respectivamente, por:

$$t_{\text{cal}_{b1}} = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}; \text{ e}$$

$$t_{\text{cal}_{b2}} = \frac{\hat{\beta}_2}{s(\hat{\beta}_2)}.$$

Se  $|t_{\text{cal}_{bw}}| \geq t_{\text{tab}_{bw}}$  (valor  $p \leq \alpha$ ), rejeita-se  $H_{0w}$ , para  $w = 1$  e  $2$ , sendo o  $t_{\text{tab}_{bw}}$  dado por:

$t_{\alpha/2, r(tg-1)-2}$ : valor da distribuição t de Student com  $r(tg - 1) - 2$  gl do resíduo da regressão que deixa uma probabilidade  $\alpha/2$  na extremidade da cauda à direita.

### 3.3.3. Modelo de regressão

O modelo de regressão linear múltipla com duas variáveis independentes, isto é, para estudar  $Y$  em função de  $X_1$  e  $X_2$  em um experimento fatorial instalado sob o DBC com  $tgr$  unidades experimentais compostas por  $t$  níveis quantitativos de  $X_1$ ,  $g$  níveis quantitativos de  $X_2$  e  $r$  blocos, sem a interação  $X_1 \times X_2$  e sem a consideração do modelo estatístico do DBC, é dado por:

$$y_{ijk} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2j} + \varepsilon'_{ijk}, \text{ em que:}$$

$y_{ijk}$ : valor observado da variável dependente  $Y$  na combinação entre os níveis quantitativos  $x_{1_i}$  ( $i = 1, 2, \dots, t$ ) e  $x_{2_j}$  ( $j = 1, 2, \dots, g$ ) e no bloco  $b_k$  ( $k = 1, 2, \dots, r$ );

$\beta_0$ : constante da regressão;

$\beta_1$  e  $\beta_2$ : coeficientes da regressão;

$\varepsilon'_{ijk} = y_{ijk} - \mu_{ij} = \varepsilon_{ijk} + \phi_{ij}$ : erro da regressão associado ao valor observado  $y_{ijk}$ ;

$\mu_{ij} = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_j}$ : média populacional da variável dependente  $Y$  na combinação entre os níveis quantitativos  $x_{1_i}$  e  $x_{2_j}$ ; e

$\phi_{ij}$ : falta de ajuste na combinação entre os níveis quantitativos  $x_{1_i}$  e  $x_{2_j}$ .

O modelo de regressão linear múltipla foi definido conforme a expressão apresentada por Bastos et al. (2015).

Assim, para os resultados do experimento com tgr pares de valores das duas variáveis independentes ( $x_{1_i}$  e  $x_{2_j}$ ) e da variável dependente ( $y_{ijk}$ ), conclui-se que os  $Y$  dependem de  $X_1$  e  $X_2$  e dos erros da regressão ( $\varepsilon'_{ijk}$ ), como segue:

$$y_{111} = \beta_0 + \beta_1 x_{1_1} + \beta_2 x_{2_1} + \varepsilon'_{111};$$

$$y_{112} = \beta_0 + \beta_1 x_{1_1} + \beta_2 x_{2_1} + \varepsilon'_{112};$$

...

$$y_{tgr} = \beta_0 + \beta_1 x_{1_t} + \beta_2 x_{2_g} + \varepsilon'_{tgr}.$$

Em termos matriciais, o modelo linear  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$  é expresso por:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ \dots \\ y_{11r} \\ y_{121} \\ y_{122} \\ \dots \\ y_{12r} \\ \dots \\ y_{tg1} \\ y_{tg2} \\ \dots \\ y_{tgr} \end{bmatrix} = \begin{bmatrix} 1 & x_{1_1} & x_{2_1} \\ 1 & x_{1_1} & x_{2_1} \\ \dots & \dots & \dots \\ 1 & x_{1_1} & x_{2_1} \\ 1 & x_{1_1} & x_{2_2} \\ 1 & x_{1_1} & x_{2_2} \\ \dots & \dots & \dots \\ 1 & x_{1_1} & x_{2_2} \\ \dots & \dots & \dots \\ 1 & x_{1_t} & x_{2_g} \\ 1 & x_{1_t} & x_{2_g} \\ \dots & \dots & \dots \\ 1 & x_{1_t} & x_{2_g} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon'_{111} \\ \varepsilon'_{112} \\ \dots \\ \varepsilon'_{11r} \\ \varepsilon'_{121} \\ \varepsilon'_{122} \\ \dots \\ \varepsilon'_{12r} \\ \dots \\ \varepsilon'_{tg1} \\ \varepsilon'_{tg2} \\ \dots \\ \varepsilon'_{tgr} \end{bmatrix}, \text{ em que:}$$

$\mathbf{y}$ : vetor  $tgr \times 1$  dos valores observados da variável dependente  $Y$ ;

$\mathbf{X}$ : matriz de incidência  $tgr \times 3$ , sendo a primeira coluna composta pela constante 1, a segunda, pelos níveis quantitativos de  $X_1$ , e a terceira, pelos níveis quantitativos de  $X_2$ ;

$\boldsymbol{\beta}$ : vetor  $3 \times 1$  dos parâmetros  $\beta_0$ ,  $\beta_1$  e  $\beta_2$ ; e

$\boldsymbol{\varepsilon}'$ : vetor  $tgr \times 1$  dos erros da regressão.

Matricialmente, o sistema de equações normais  $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$  é dado por:

$$\begin{bmatrix} tgr & gr \sum_{i=1}^t x_{1i} & tr \sum_{j=1}^g x_{2j} \\ gr \sum_{i=1}^t x_{1i} & gr \sum_{i=1}^t x_{1i}^2 & r \sum_{i=1}^t \sum_{j=1}^g x_{1i} x_{2j} \\ tr \sum_{j=1}^g x_{2j} & r \sum_{i=1}^t \sum_{j=1}^g x_{1i} x_{2j} & tr \sum_{j=1}^g x_{2j}^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r y_{ijk} \\ \sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r x_{1i} y_{ijk} \\ \sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r x_{2j} y_{ijk} \end{bmatrix}$$

Assim, após a resolução do vetor  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , apresenta as estimativas dos parâmetros  $\beta_0$ ,  $\beta_1$  e  $\beta_2$ , como segue:

$$\hat{y}_{ijk} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2j}, \text{ em que:}$$

$\hat{y}_{ijk}$ : valor ajustado da variável dependente  $Y$  na combinação entre os níveis quantitativos  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e  $x_{2j}$  ( $j = 1, 2, \dots, g$ ) e no bloco  $b_k$  ( $k = 1, 2, \dots, r$ ).

Do mesmo modo, em termos de média, os estimadores de mínimos quadrados  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  e  $\hat{\beta}_2$  são não tendenciosos, uma vez que:

$$E(\hat{\beta}_0) = \beta_0;$$

$$E(\hat{\beta}_1) = \beta_1; \text{ e}$$

$$E(\hat{\beta}_2) = \beta_2.$$

Já as estimativas das variâncias dos estimadores de mínimos quadrados  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  e  $\hat{\beta}_2$  são obtidas na diagonal da seguinte matriz:

$$\hat{V}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} s^2(\hat{\beta}_0) & s(\hat{\beta}_0, \hat{\beta}_1) & s(\hat{\beta}_0, \hat{\beta}_2) \\ s(\hat{\beta}_0, \hat{\beta}_1) & s^2(\hat{\beta}_1) & s(\hat{\beta}_1, \hat{\beta}_2) \\ s(\hat{\beta}_0, \hat{\beta}_2) & s(\hat{\beta}_1, \hat{\beta}_2) & s^2(\hat{\beta}_2) \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \text{QMResReg, em que:}$$

QMResReg: quadrado médio do resíduo da regressão (Tabela 3.2).

As estimativas dos desvios-padrão dos estimadores  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  e  $\hat{\beta}_2$ , fornecidas por  $s(\hat{\beta}_0)$ ,  $s(\hat{\beta}_1)$  e  $s(\hat{\beta}_2)$ , são obtidas extraíndo-se a raiz quadrada das respectivas estimativas das variâncias.

### 3.3.4. Delineamento composto central

Além do fatorial completo  $t \times g$  utilizado para obtenção de um experimento com duas variáveis independentes  $X_1$  e  $X_2$ , conforme já mencionado, pode-se, também, utilizar fatoriais incompletos, isto é, que proporcionam economia no número de combinações (tratamentos).

O delineamento composto central (DCC) é um delineamento simétrico e de segunda ordem, constituído por pontos cúbicos ( $\pm 1$ ), axiais ( $\pm \alpha$ ) e central (0), de forma bastante flexível. Ele é um delineamento simples e econômico, muito utilizado dentro da metodologia de superfície de resposta, que foi introduzido por Box e Wilson (1951) no estudo de superfícies de respostas como uma evolução do delineamento  $3^k$ , sendo  $k$  o número de variáveis independentes (Tabela 3.3).

O DCC possui características interessantes para a busca do ponto em que se dá a resposta ótima, como por exemplo, propiciar um menor número de tratamentos em relação aos fatoriais completos e ser realizado sequencialmente, de forma a caminhar no sentido da otimização do processo, isto é, através das execuções de experimentos menores.

Tabela 3.3 - Combinações entre os níveis codificados planejadas pelo DCC com duas variáveis independentes ( $X_1$  e  $X_2$ )

Tratamento	$x_{1c}$	$x_{2c}$
1	-1	-1
2	1	-1
3	-1	1
4	1	1
5	$-\alpha$	0
6	$\alpha$	0
7	0	$-\alpha$
8	0	$\alpha$
9	0	0

Segundo Haaland (1989), o número de repetições do ponto central tem por finalidade fornecer estimativas do erro experimental e de contribuir, também, para as estimativas dos termos quadráticos. Já os pontos centrais melhoram as estimativas dos efeitos quadráticos e possibilitam graus de liberdade adicionais para o erro experimental. Além disso, os pontos

centrais permitem informações extras quanto ao centro da região experimental, onde, possivelmente, os melhores valores da resposta estão frequentemente localizados. Os pontos cúbicos são os que contribuem para as estimativas das interações entre as variáveis independentes.

Por outro lado, os pontos axiais são úteis para as estimativas dos termos quadráticos e, portanto, pela localização do ponto ótimo, caso ele ocorra na região estudada. É o valor do ponto axial ( $\alpha$ ) que particulariza o DCC, pois ele pode ser escolhido para torná-lo ortogonal e/ou para dar a ele a propriedade de ser rotacional. Isso significa que há várias possibilidades para a sua escolha (RODRIGUES; IEMMA, 2005).

Um DCC é considerado rotacional, quando as variâncias das predições da resposta dependerem apenas da distância em relação ao ponto central, isto é, quando a precisão da resposta prevista for a mesma em todos os pontos situados numa dada (hiper) esfera com o centro situado no próprio centro do delineamento. Assim, a rotacionalidade propicia variâncias idênticas para pontos situados à mesma distância do centro, em qualquer direção.

De acordo com Box e Draper (1987), Barros Neto et al. (2003) e Myers et al. (2009), o valor mais usual de  $\alpha$  é aquele que torna o DCC rotacional (DCCR). Ele melhora a qualidade de estimação, por proporcionar a mesma variação da média estimada dentro de cada combinação entre os níveis das variáveis independentes estudadas, visto que todos os pontos são equidistantes ao central em qualquer direção. No caso em que o DCC for completo para duas variáveis independentes ( $X_1$  e  $X_2$ ), tem se:

$$\alpha = \sqrt[4]{2^2} = 1,4142.$$

Segundo Barros Neto et al. (2003) e Ribeiro Júnior (2012), quando o valor escolhido for  $\alpha = 1$ , os pontos axiais estarão localizados sobre o cubo e o delineamento será denominado de DCC face centrada (DCCFC).

Se a escolha for para  $\alpha > 1$ , além de ser possível obter o DCC ortogonal e/ou rotacional, o seu aumento implicará em pontos cúbicos cada vez mais próximos do ponto central. Caso contrário, para  $0 < \alpha < 1$ , a sua diminuição fará com que os pontos axiais tenda em direção ao ponto central.

### 3.3.5. Fatorial $2^2 + 1$

No fatorial mais simples  $2^2$ , no qual há dois níveis das variáveis independentes  $X_1$  e  $X_2$ , são avaliados apenas quatro tratamentos (YATES, 1937). Assim, o fatorial com tratamento adicional ( $2^2 + 1$ ) é um caso particular do fatorial completo  $2^2$ . Nesse caso, o tratamento adicional corresponde a uma combinação entre os níveis de  $X_1$  e  $X_2$  que não faz parte da estrutura fatorial, ou seja, que têm níveis que não combinam com os demais níveis de  $X_1$  e  $X_2$  (Tabela 3.4). Por isso, Zeviane (2011) classifica os experimentos fatoriais com um ou mais tratamentos adicionais como fatoriais incompletos.

Tabela 3.4 - Combinações entre os níveis codificados planejadas pelo fatorial  $2^2 + 1$

Tratamento	$x_{1c}$	$x_{2c}$
1	-1	-1
2	1	-1
3	-1	1
4	1	1
5	0	0

A inclusão de um ou mais tratamentos adicionais em uma estrutura fatorial é muito comum e pode ser importante para servir como termo de referência para avaliação dos efeitos dos demais tratamentos e, também, para obtenção de informações complementares. Além disso, podem, ainda, serem incluídos com o intuito de diminuir o número de tratamentos. Para variáveis independentes com níveis quantitativos, o tratamento adicional é, geralmente, constituído pelo ponto central.

Um exemplo de utilização do experimento fatorial com um tratamento adicional foi feito por Oliveira et al. (1999). Eles avaliaram as influências das aplicações de ureia e vinhaça sobre a degradação da lignocelulose e a liberação de nutrientes da palha da cana-de-açúcar. Os tratamentos foram constituídos de duas formas de aplicação da ureia (sobre a palha da cana recém colhida e no solo à profundidade de 15 cm) combinadas com duas fontes de potássio ( $KC_1$  e vinhaça) e, ainda, de um tratamento adicional (testemunha) formado pela palhada recém colhida sem aplicação de ureia. Isso significa que o experimento foi delineado conforme o fatorial  $2^2 + 1$ .

### 3.4. Pressuposições

Ao serem estabelecidos os modelos de regressão lineares simples e múltiplo, pressupõe-se que:

a) a média dos erros da regressão é nula, isto é,  $E(\varepsilon') = 0$ ;

b) para todos os níveis de  $X_1$  (regressão linear simples) ou para todas as combinações entre os níveis de  $X_1, X_2, \dots, X_k$  (regressão linear múltipla), as variâncias dos erros da regressão devem ser homogêneas, o que significa dizer que elas devem ser homocedásticas, isto é, que  $V(\varepsilon') = \sigma_{\varepsilon'}^2$ ;

c) os erros da regressão devem ser independentes; e

d) os erros da regressão devem ter distribuição normal de probabilidades.

Combinando-se as três últimas pressuposições, tem-se que  $\varepsilon' \sim \text{NIID } (0, \sigma_{\varepsilon'}^2)$ , em que NIID significa normal, independente e identicamente distribuído. De forma mais resumida, pode-se adotar que  $\varepsilon' \sim N(0, \sigma_{\varepsilon'}^2)$ .

## 4. MATERIAL E MÉTODOS

### 4.1. Regressão linear simples

#### 4.1.1. Parâmetros da regressão

O modelo de regressão linear simples que representou a relação funcional entre a variável dependente (Y) e a variável independente ( $X_1$ ) foi dado por:

$$y_{ij} = 1.000 + 10x_{1i} + \varepsilon'_{ij}, \text{ para } 0 \leq X_1 \leq 100, \text{ em que:}$$

$y_{ij}$ : valor observado da variável dependente Y no nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e no bloco

$b_j$  ( $j = 1, 2, \dots, r$ );

$\beta_0 = 1.000$ : constante da regressão;

$\beta_1 = 10$ : coeficiente da regressão;

$\varepsilon'_{ij}$ : erro da regressão associado ao valor observado  $y_{ij}$ ;

$\mu_i = 1.000 + 10x_{1i}$ : média populacional da variável dependente Y no nível quantitativo  $x_{1i}$ ; e

$\mu = 1.500$ : média populacional geral da variável dependente Y.

Os parâmetros da regressão ( $\beta_0$  e  $\beta_1$ ) foram definidos com base na equação de regressão linear simples ajustada às seguintes variáveis avaliadas em um experimento com soja: início da maturidade (Y) e dose de aplicação do nitrogênio ( $X_1$ ). Este experimento consistiu em aplicar o nitrogênio durante a fase reprodutiva da soja, entre os estágios  $R_1$  (início do florescimento) e  $R_6$  (grãos completamente cheios preenchendo a cavidade do legume), e avaliar o início da maturidade no estágio  $R_7$  (BAHRY et al., 2013).

#### 4.1.2. Simulação dos dados

Para obtenção dos resíduos da regressão ( $e'$ ), foram realizadas 1.000 simulações, para cada cenário analisado, de acordo com a distribuição normal com média populacional igual a zero e desvio-padrão populacional  $\sigma_{e'}$ , sendo:

$e'_{ij}$ : resíduo da regressão associado ao valor observado  $y_{ij}$  ( $i = 1, 2, \dots, t$  e  $j = 1, 2, \dots, r$ ).

Num primeiro momento, para as realizações das simulações, os valores de  $\sigma_{\varepsilon'}$  foram definidos para fornecerem coeficientes de variação residuais ( $CV_{\varepsilon'}$ ) iguais a 10%, 20% e 30%, conforme a seguinte expressão:

$$CV_{\varepsilon'} = 100 \times \frac{\sigma_{\varepsilon'}}{\mu} = 100 \times \frac{\sigma_{\varepsilon'}}{1.500}.$$

Para as realizações das diferentes simulações, foram adotados valores de  $\sigma_{\varepsilon'}$  iguais a 150, 300 e 450, respectivamente. Consequentemente, têm-se as seguintes distribuições normais:

$$\varepsilon'_{ij} \sim N(\mu_{\varepsilon'} = 0; \sigma_{\varepsilon'}^2 = 150^2);$$

$$\varepsilon'_{ij} \sim N(\mu_{\varepsilon'} = 0; \sigma_{\varepsilon'}^2 = 300^2); \text{ e}$$

$$\varepsilon'_{ij} \sim N(\mu_{\varepsilon'} = 0; \sigma_{\varepsilon'}^2 = 450^2).$$

#### 4.1.3. Delineamento em blocos casualizados

Para fins de comparação, foram gerados 25 experimentos instalados sob o DBC conforme 25 combinações entre o número de tratamentos (níveis quantitativos de  $X_1$  entre zero e 100) (t) e o número de blocos (r), a fim de fornecerem os mesmos números de unidades experimentais ( $n = tr$ ) iguais a 12, 16, 20, 24, 28 e 32.

Já os efeitos de blocos, tomando-se como referências os experimentos instalados sob o DBC com  $CV_{\varepsilon'} = 30\%$ , foram definidos a fim de fornecerem, aproximadamente, as mesmas somas de quadrados de blocos (SQBI) que promovessem significâncias para os próprios efeitos ( $f_{\text{cal}_{\text{BI}}} \geq f_{\text{tab}_{\text{BI}}}$ ) em todos os experimentos com o mesmo valor de n. Para isso, foi considerado  $\alpha = 0,05$ .

Nesse estudo, foram gerados quatro experimentos instalados sob o DBC para  $n = 12$  (Tabela 4.1), três para  $n = 16$  (Tabela 4.2), quatro para  $n = 20$  (Tabela 4.3), seis para  $n = 24$  (Tabela 4.4), quatro para  $n = 28$  (Tabela 4.5) e quatro para  $n = 32$  (Tabela 4.6), sendo:

$x_{1,i}$ : nível quantitativo da variável independente  $X_1$  ( $i = 1, 2, \dots, t$ ); e

$\omega_j$ : efeito do bloco  $b_j$  ( $j = 1, 2, \dots, r$ ).

Tabela 4.1 - Níveis quantitativos e efeitos de blocos dos quatro experimentos instalados sob o DBC com  $n = 12$

t = 2 e r = 6		t = 3 e r = 4		t = 4 e r = 3		t = 6 e r = 2	
$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$
0	-500	0	-432	0	-418	0	-342
100	-300	50	-216	33,33	0	20	342
-	-100	100	216	66,67	418	40	-
-	100	-	432	100	-	60	-
-	300	-	-	-	-	80	-
-	500	-	-	-	-	100	-

Tabela 4.2 - Níveis quantitativos e efeitos de blocos dos três experimentos instalados sob o DBC com  $n = 16$

t = 2 e r = 8		t = 4 e r = 4		t = 8 e r = 2	
$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$
0	-400	0	-346	0	-274
100	-300	33,33	-173	14,29	274
-	-200	66,67	173	28,57	-
-	-100	100	346	42,86	-
-	100	-	-	57,14	-
-	200	-	-	71,43	-
-	300	-	-	85,71	-
-	400	-	-	100	-



Tabela 4.5 - Níveis quantitativos e efeitos de blocos dos quatro experimentos instalados sob o DBC com  $n = 28$

t = 2 e r = 14		t = 4 e r = 7		t = 7 e r = 4		t = 14 e r = 2	
$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$
0	-321	0	-276	0	-233	0	-184
100	-261	33,33	-184	16,67	-116	7,69	184
-	-201	66,67	-92	33,33	116	15,38	-
-	-141	100	0	50	233	23,08	-
-	-81	-	92	66,67	-	30,77	-
-	-21	-	184	83,33	-	38,46	-
-	0	-	276	100	-	46,15	-
-	0	-	-	-	-	53,85	-
-	21	-	-	-	-	61,54	-
-	81	-	-	-	-	69,23	-
-	141	-	-	-	-	76,92	-
-	201	-	-	-	-	84,62	-
-	261	-	-	-	-	92,31	-
-	321	-	-	-	-	100	-

Tabela 4.6 - Níveis quantitativos e efeitos de blocos dos quatro experimentos instalados sob o DBC com  $n = 32$

t = 2 e r = 16		t = 4 e r = 8		t = 8 e r = 4		t = 16 e r = 2	
$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$	$x_{1i}$	$\omega_j$
0	-260	0	-247	0	-214	0	-170
100	-230	33,33	-185	14,29	-108	6,67	170
-	-200	66,67	-125	28,57	108	13,33	-
-	-170	100	-65	42,86	214	20	-
-	-140	-	65	57,14	-	26,67	-
-	-110	-	125	71,43	-	33,33	-
-	-80	-	185	85,71	-	40	-
-	-50	-	247	100	-	46,67	-
-	50	-	-	-	-	53,33	-
-	80	-	-	-	-	60	-
-	110	-	-	-	-	66,67	-
-	140	-	-	-	-	73,33	-
-	170	-	-	-	-	80	-
-	200	-	-	-	-	86,67	-
-	230	-	-	-	-	93,33	-
-	260	-	-	-	-	100	-

Desse modo, para cada uma das 18 combinações entre os valores do  $CV_{\varepsilon'}$  (10%, 20% e 30%) e do  $n$  (12, 16, 20, 24, 28 e 32), foram realizadas 1.000 simulações de acordo com as respectivas distribuições normais ( $\mu_{\varepsilon'} = 0$  e  $\sigma_{\varepsilon'}$ ), a fim de gerar os  $n$  (tr) resíduos da regressão.

Posteriormente, os valores observados da variável dependente  $Y$  em cada um dos 25 experimentos balanceados e instalados sob o DBC, foram obtidos por:

$$y_{ij} = 1.000 + 10x_{1i} + \omega_j + e'_{ij}, \text{ em que:}$$

$y_{ij}$ : valor observado da variável dependente  $Y$  no nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e no bloco  $b_j$  ( $j = 1, 2, \dots, r$ ).

Foram gerados, no total, 75 ( $3 \times 25$ ) diferentes conjuntos de dados para o estudo da análise de regressão linear simples e, para cada um deles, realizadas 1.000 simulações.

Para cada um dos 75.000 conjuntos de dados, foi ajustado um modelo de regressão linear de primeiro grau, como segue:

$$\hat{Y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i}, \text{ em que:}$$

$\hat{Y}_{ij}$ : valor ajustado da variável dependente Y no nível quantitativo  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e no bloco  $b_j$  ( $j = 1, 2, \dots, r$ ).

Logo depois, foram realizadas as análises de variância da regressão com o teste da falta de ajuste sob o DBC de um experimento balanceado (Tabela 4.8).

Tabela 4.8 - Análise de variância da regressão com o teste para a falta de ajuste

FV	GL	SQ	QM	F
Bloco	$r - 1$	SQBI	–	–
Regressão	1	SQReg	SQReg/1	QMReg/QMResReg
ResReg	$r(t - 1) - 1$	SQResReg	$SQResReg/[r(t - 1) - 1]$	
Falta de Ajuste	$t - 2$	SQFA	$SQFA/(t - 2)$	QMFA/QMRes
Resíduo	$(t - 1)(r - 1)$	SQRes	$SQRes/[(t - 1)(r - 1)]$	

#### 4.1.4. Medidas avaliadas

Para comparar, dentro de cada valor de n (12, 16, 20, 24, 28 e 32), os efeitos do  $CV_{\epsilon}$ , em percentual (cv = 10%, 20% e 30%) e do número de níveis quantitativos de  $X_1$  ( $t = 2, 3, 4$  e 6, para n = 12, t = 2, 4 e 8, para n = 16, t = 2, 4, 5 e 10, para n = 20, t = 2, 3, 4, 6, 8 e 12, para n = 24, t = 2, 4, 7 e 14, para n = 28 e t = 2, 4, 8 e 16, para n = 32), com base nas 1.000 simulações, foram analisadas as cinco variáveis definidas a seguir:

$$EPMA_{\beta_0} = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{0s} - \beta_0}{\beta_0} \right| \times 100 = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{0s} - 1.000}{1.000} \right| \times 100;$$

$$EPMA_{\beta_1} = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{1s} - \beta_1}{\beta_1} \right| \times 100 = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{1s} - 10}{10} \right| \times 100;$$

$$EPMA_{\mu} = \frac{1}{1.000} \sum_{s=1}^{1.000} f_s;$$

$$R = \frac{1}{1.000} \sum_{s=1}^{1.000} \frac{SQReg_s}{SQReg_s + SQResReg_s}; e$$

$$ER = \frac{1}{1.000} \sum_{s=1}^{1.000} \frac{(r-1)QMBl_s + r(t-1)QMResReg_s}{(tr-1)QMResReg_s}, \text{ em que:}$$

$f_s = \frac{1}{16} \sum_{i=1}^{16} \left| \frac{\hat{y}_{ij} - \mu_i}{\mu_i} \right| \times 100$ , para  $x_{1_i} = 0; 6,67; 13,33; 20; 26,67; 33,33; 40; 46,67; 53,33; 60; 66,67; 73,33; 80; 86,67; 93,33$ ; e 100.

Os EPMA's (erro percentual médio absoluto) mostram as diferenças absolutas entre os parâmetros e as estimativas obtidas pelos respectivos modelos ajustados de regressão linear de primeiro grau. Para uma perfeita análise, esperar-se-ia que todos eles fossem iguais a zero. E para as medidas R e ER, quanto maiores forem os seus valores, melhores foram o ajuste do modelo de regressão linear de primeiro grau e a eficiência do DBC, respectivamente.

Como já mencionado, para  $n = 12$  (Tabela 4.1), serão analisados os efeitos dos três valores do  $CV_{\varepsilon'}$  e de quatro níveis quantitativos ( $t = 2, 3, 4$  e  $6$ ). Para  $n = 16$  (Tabela 4.2), os mesmos valores do  $CV_{\varepsilon'}$  foram combinados com três números de níveis quantitativos ( $t = 2, 4$  e  $8$ ). Para  $n = 20$  (Tabela 4.3), quatro níveis quantitativos ( $t = 2, 4, 5$  e  $10$ ). Para  $n = 24$  (Tabela 4.4), com seis níveis quantitativos ( $t = 2, 3, 4, 6, 8$  e  $12$ ). Para  $n = 28$  (Tabela 4.5), com quatro níveis quantitativos ( $t = 2, 4, 7$  e  $14$ ). E para  $n = 32$  (Tabela 4.6), quatro níveis quantitativos ( $t = 2, 4, 8$  e  $16$ ).

Isso significa que foram gerados seis experimentos fatoriais ( $3 \times 4, 3 \times 3, 3 \times 4, 3 \times 6, 3 \times 4$  e  $3 \times 4$ ) sem repetições por combinação entre os níveis dos dois fatores ( $CV_{\varepsilon'}$  e  $t$ ), com base nas médias das 1.000 simulações e de acordo com o delineamento inteiramente casualizado (DIC).

Para cada uma das cinco medidas avaliadas ( $EPMA_{\beta_0}, EPMA_{\beta_1}, EPMA_{\mu}, R$  e  $ER$ ) em cada um dos 75 diferentes conjuntos de dados, separadamente, foi realizada uma análise de superfície de resposta para verificar os efeitos do número de níveis quantitativos ( $t$ ) e do  $CV_{\varepsilon'}$  ( $cv$ ) para cada número de unidades experimentais ( $n$ ), separadamente, cujo maior modelo adotado foi definido por:

$$y_{ij} = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \beta_3 b_j + \beta_4 b_j^2 + \beta_5 a_i b_j + \varepsilon'_{ij}, \text{ em que:}$$

$y_{ij}$ : valor observado da medida avaliada na combinação entre os níveis relacionados aos números de níveis quantitativos ( $a_i$ ) [ $2, 3, 4$  e  $6$  ( $n = 12$ ),  $2, 4$  e  $8$  ( $n = 16$ ),  $2, 4, 5$  e  $10$  ( $n = 20$ ),  $2, 3, 4, 6, 8$  e  $12$  ( $n = 24$ ),  $2, 4, 7$  e  $14$  ( $n = 28$ ) e  $2, 4, 8$  e  $16$  ( $n = 32$ )] e aos valores do  $CV_{\varepsilon'}$  ( $b_j$ ) ( $10, 20$  e  $30$ );

$\beta_0$ : constante da regressão;

$\beta_1, \beta_2, \beta_3, \beta_4$  e  $\beta_5$ : coeficientes da regressão; e

$\varepsilon'_{ij} \sim N(0, \sigma_{\varepsilon'}^2)$ .

Posteriormente, para o ajuste do melhor modelo, foram retirados, um de cada vez, e começando pelo mais complexo de interpretação, pelo menos um efeito não significativo, caso haja, de acordo com o teste t de Student a 5% de significância. Caso sejam de mesma complexidade, foi retirado o que possuiu o maior valor p, desde que não significativos. No entanto, os efeitos não significativos que apresentarem menor hierarquia em relação aos respectivos efeitos significativos foram mantidos no modelo.

As análises estatísticas realizadas dentro de cada valor de n (12, 16, 20, 24, 28 e 32) tiveram como objetivo verificar se, para diferentes experimentos instalados sob o DBC com diferentes precisões, foi melhor avaliar menos níveis quantitativos com mais repetições ou mais níveis quantitativos com menos repetições, considerando-se o mesmo número de unidades experimentais (n) em uma análise de regressão linear simples.

Todas as simulações e análises estatísticas referentes às avaliações do modelo de regressão linear simples foram realizadas no R versão 4.0.2 (R CORE TEAM, 2020).

## 4.2. Regressão linear múltipla

### 4.2.1. Parâmetros da regressão

O modelo de regressão linear múltipla que representou a relação funcional entre a variável dependente (Y) e as duas variáveis independentes ( $X_1$  e  $X_2$ ) foi dado por:

$$y_{ijk} = 1.000 + 10x_{1i} + 10x_{2j} + \varepsilon'_{ijk}, \text{ para } 0 \leq X_1 \leq 100 \text{ e } 0 \leq X_2 \leq 100, \text{ em que:}$$

$y_{ijk}$ : valor observado da variável dependente Y na combinação entre os níveis quantitativos  $x_{1i}$  ( $i = 1, 2, \dots, t$ ) e  $x_{2j}$  ( $j = 1, 2, \dots, g$ ) e no bloco  $b_k$  ( $k = 1, 2, \dots, r$ );

$\beta_0 = 1.000$ : constante da regressão;

$\beta_1 = \beta_2 = 10$ : coeficientes da regressão;

$\varepsilon'_{ijk}$ : erro da regressão associado ao valor observado  $y_{ijk}$ ;

$\mu_{ij} = 1.000 + 10x_{1i} + 10x_{2j}$ : média populacional da variável dependente Y na combinação entre os níveis quantitativos  $x_{1i}$  e  $x_{2j}$ ; e

$\mu = 2.000$ : média populacional geral da variável dependente Y.

### 4.2.2. Simulação dos dados

Para obtenção dos resíduos da regressão ( $e'$ ), foram realizadas 1.000 simulações de acordo com a distribuição normal com média populacional igual a zero e desvio-padrão populacional  $\sigma_{e'}$ , sendo:

$e'_{ijk}$ : resíduo da regressão associado ao valor observado  $y_{ijk}$  ( $i = 1, 2, \dots, j = 1, 2, \dots$  e  $k = 1, 2, \dots, r$ ).

Num primeiro momento, para as realizações das simulações, o valor de  $\sigma_{e'}$  foi definido para fornecer um coeficiente de variação residual ( $CV_{e'}$ ) igual a 20%, conforme a seguinte expressão:

$$CV_{e'} = 100 \times \frac{\sigma_{e'}}{\mu} = 100 \times \frac{\sigma_{e'}}{2.000}.$$

Portanto, para as realizações das 1.000 simulações, foi adotado o valor de  $\sigma_{e'}$  igual a 400. Conseqüentemente, tem-se a seguinte distribuição normal:

$$e'_{ijk} \sim N (\mu_{e'} = 0; \sigma_{e'}^2 = 400^2).$$

### 4.2.3. Delineamento em blocos casualizados

Para fins de comparação, foram gerados 15 experimentos instalados sob o DBC conforme 15 combinações entre o número de tratamentos (combinações entre os níveis quantitativos de  $X_1$  e  $X_2$  entre zero e 100) ( $n_{\text{trat}}$ ) e o número de blocos ( $r$ ), a fim de fornecerem os mesmos números de unidades experimentais ( $n$ ) iguais a 18, 24, 30 e 36.

Nesse estudo, foram gerados três experimentos instalados sob o DBC para  $n = 18$  (Tabela 4.9), três para  $n = 24$  (Tabela 4.10), três para  $n = 30$  (Tabela 4.11) e seis para  $n = 36$  (Tabela 4.12), sendo:

$x_{1i}$ : nível quantitativo da variável independente  $X_1$  ( $i = 1, 2, \dots$ );

$x_{2j}$ : nível quantitativo da variável independente  $X_2$  ( $j = 1, 2, \dots$ ); e

$\omega_k$ : efeito do bloco  $b_k$  ( $k = 1, 2, \dots, r$ ).

Tabela 4.9 - Combinações entre os níveis quantitativos e efeitos de blocos dos três experimentos fatoriais instalados sob o DBC com  $n = 18$

$t = 6 (2 \times 3) e r = 3$		$t = 9 (DCCR) e r = 2$		$t = 9 (DCCFC) e r = 2$	
$x_{1_i}   x_{2_j}$	$\omega_k$	$x_{1_i}   x_{2_j}$	$\omega_k$	$x_{1_i}   x_{2_j}$	$\omega_k$
0   0	-645	14,64   14,64	-735	0   0	-735
100   0	0	85,36   14,64	735	100   0	735
0   50	645	14,64   85,36	-	0   100	-
100   50	-	85,36   85,36	-	100   100	-
0   100	-	0   50	-	0   50	-
100   100	-	100   50	-	100   50	-
-	-	50   0	-	50   0	-
-	-	50   100	-	50   100	-
-	-	50   50	-	50   50	-

Tabela 4.10 - Combinações entre os níveis quantitativos e efeitos de blocos dos três experimentos fatoriais instalados sob o DBC com  $n = 24$

$t = 4 (2 \times 2) e r = 6$		$t = 6 (2 \times 3) e r = 4$		$t = 12 (3 \times 4) e r = 2$	
$x_{1_i}   x_{2_j}$	$\omega_k$	$x_{1_i}   x_{2_j}$	$\omega_k$	$x_{1_i}   x_{2_j}$	$\omega_k$
0   0	-470	0   0	-500	0   0	-700
100   0	-250	100   0	-290	50   0	700
0   100	-100	0   50	290	100   0	-
100   100	100	100   50	500	0   33,33	-
-	250	0   100	-	50   33,33	-
-	470	100   100	-	100   33,33	-
-	-	-	-	0   66,67	-
-	-	-	-	50   66,67	-
-	-	-	-	100   66,67	-
-	-	-	-	0   100	-
-	-	-	-	50   100	-
-	-	-	-	100   100	-

Tabela 4.11 - Combinações entre os níveis quantitativos e efeitos de blocos dos três experimentos fatoriais instalados sob o DBC com  $n = 30$

$t = 5 (2^2 + 1) e r = 6$		$t = 6 (2 \times 3) e r = 5$		$t = 15 (3 \times 5) e r = 2$	
$x_{1_i}   x_{2_j}$	$\omega_k$	$x_{1_i}   x_{2_j}$	$\omega_k$	$x_{1_i}   x_{2_j}$	$\omega_k$
0   0	-450	0   0	-500	0   0	-685
100   0	-250	100   0	-200	50   0	685
0   100	-100	0   50	0	100   0	-
100   100	100	100   50	200	0   25	-
50   50	250	0   100	500	50   25	-
-	450	100   100	-	100   25	-
-	-	-	-	0   50	-
-	-	-	-	50   50	-
-	-	-	-	100   50	-
-	-	-	-	0   75	-
-	-	-	-	50   75	-
-	-	-	-	100   75	-
-	-	-	-	0   100	-
-	-	-	-	50   100	-
-	-	-	-	100   100	-

Tabela 4.12 - Combinações entre os níveis quantitativos e efeitos de blocos dos seis experimentos fatoriais instalados sob o DBC com  $n = 36$

t = 4 (2 × 2) e r = 9			t = 6 (2 × 3) e r = 6			t = 9 (DCCR) e r = 4			t = 9 (DCCFC) e r = 4			t = 12 (3 × 4) e r = 3			t = 18 (3 × 6) e r = 2		
$x_{1i}   x_{2j}$	$\omega_k$		$x_{1i}   x_{2j}$	$\omega_k$		$x_{1i}   x_{2j}$	$\omega_k$		$x_{1i}   x_{2j}$	$\omega_k$		$x_{1i}   x_{2j}$	$\omega_k$		$x_{1i}   x_{2j}$	$\omega_k$	
0   0	-380		0   0	-365		14,64   14,64	-500		0   0	-500		0   0	-590		0   0	-670	
100   0	-250		100   0	-300		85,36   14,64	-235		100   0	-235		50   0	0		50   0	670	
0   100	-150		0   50	-200		14,64   85,36	235		0   100	235		100   0	590		100   0	-	
100   100	-100		100   50	200		85,36   85,36	500		100   100	500		0   33,33	-		0   20	-	
-	0		0   100	300		0   50	-		0   50	-		50   33,33	-		50   20	-	
-	100		100   100	365		100   50	-		100   50	-		100   33,33	-		100   20	-	
-	150		-	-		50   0	-		50   0	-		0   66,67	-		0   40	-	
-	250		-	-		50   100	-		50   100	-		50   66,67	-		50   40	-	
-	380		-	-		50   50	-		50   50	-		100   66,67	-		100   40	-	
-	-		-	-		-	-		-	-		0   100	-		0   60	-	
-	-		-	-		-	-		-	-		50   100	-		50   60	-	
-	-		-	-		-	-		-	-		100   100	-		100   60	-	
-	-		-	-		-	-		-	-		-	-		0   80	-	
-	-		-	-		-	-		-	-		-	-		50   80	-	
-	-		-	-		-	-		-	-		-	-		100   80	-	
-	-		-	-		-	-		-	-		-	-		0   100	-	
-	-		-	-		-	-		-	-		-	-		50   100	-	
-	-		-	-		-	-		-	-		-	-		100   100	-	

Como se pode observar para  $n = 18$  (Tabela 4.9) e para  $n = 36$  (Tabela 4.12), o DCCFC é um delineamento fatorial  $3 \times 3$ . Portanto, a partir de agora, ele passou a receber a última denominação.

Desse modo, para cada um dos quatro valores de  $n$  (18, 24, 30 e 36), foram realizadas 1.000 simulações de acordo com a respectiva distribuição normal ( $\mu_{\epsilon'} = 0$  e  $\sigma_{\epsilon'} = 400$ ), a fim de gerar os  $n$  e respectivos resíduos da regressão.

Posteriormente, os valores observados da variável dependente  $Y$  em cada um dos 15 experimentos balanceados e instalados sob o DBC, foram obtidos por:

$$y_{ijk} = 1.000 + 10x_{1i} + 10x_{2j} + \omega_k + e'_{ijk}, \text{ em que:}$$

$y_{ijk}$ : valor observado da variável dependente  $Y$  na combinação entre os níveis quantitativos  $x_{1i}$  ( $i = 1, 2, \dots$ ) e  $x_{2j}$  ( $j = 1, 2, \dots$ ) e no bloco  $b_k$  ( $k = 1, 2, \dots, r$ ).

Foram gerados, no total, 15 diferentes conjuntos de dados para o estudo da análise de regressão linear múltipla e, para cada um deles, realizadas 1.000 simulações.

Para cada um dos 15.000 conjuntos de dados, foi ajustado um modelo de regressão linear de primeira ordem, como segue:

$$\hat{Y}_{ijk} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2j}, \text{ em que:}$$

$\hat{Y}_{ijk}$ : valor ajustado da variável dependente Y na combinação entre os níveis quantitativos  $x_{1i}$  ( $i = 1, 2, \dots$ ) e  $x_{2j}$  ( $j = 1, 2, \dots$ ) e no bloco  $b_k$  ( $k = 1, 2, \dots, r$ ).

Logo depois, foram realizadas as análises de variância da regressão com o teste da falta de ajuste sob o DBC de um experimento balanceado (Tabela 4.13).

Tabela 4.13 - Análise de variância da regressão com o teste para a falta de ajuste

FV	GL	SQ	QM	F
Bloco	$r - 1$	SQBI	–	–
Regressão	2	SQReg	SQReg/2	QMReg/QMResReg
ResReg	$n - r - 2$	SQResReg	SQResReg/( $n - r - 2$ )	
Falta de Ajuste	$n_{\text{trat}} - 3$	SQFA	SQFA/( $n_{\text{trat}} - 3$ )	QMFA/QMRes
Resíduo	$n - r - n_{\text{trat}} + 1$	SQRes	SQRes/( $n - r - n_{\text{trat}} + 1$ )	

#### 4.2.4. Medidas avaliadas

Para comparar, dentro de cada valor de  $n$  (18, 24, 30 e 36), os diferentes experimentos, com base nas 1.000 simulações, foram analisadas as quatro variáveis definidas a seguir:

$$EPMA_{\beta_0} = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{0s} - \beta_0}{\beta_0} \right| \times 100 = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{0s} - 1.000}{1.000} \right| \times 100;$$

$$EPMA_{\beta_1} = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{1s} - \beta_1}{\beta_1} \right| \times 100 = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{1s} - 10}{10} \right| \times 100;$$

$$EPMA_{\beta_2} = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{2s} - \beta_2}{\beta_2} \right| \times 100 = \frac{1}{1.000} \sum_{s=1}^{1.000} \left| \frac{\hat{\beta}_{2s} - 10}{10} \right| \times 100; \text{ e}$$

$$R = \frac{1}{1.000} \sum_{s=1}^{1.000} \frac{SQReg_s}{SQReg_s + SQResReg_s}.$$

Neste estudo, foram comparados os seguintes delineamentos fatoriais:

$2 \times 3$ ,  $3 \times 3$  (DCCFC) e DCCR, para  $n = 18$  (Tabela 4.9);

$2 \times 2$ ,  $2 \times 3$  e  $3 \times 4$ , para  $n = 24$  (Tabela 4.10);

$2^2 + 1$ ,  $2 \times 3$  e  $3 \times 5$ , para  $n = 30$  (Tabela 4.11); e

$2 \times 2$ ,  $2 \times 3$ ,  $3 \times 3$  (DCCFC),  $3 \times 4$ ,  $3 \times 6$  e DCCR, para  $n = 36$  (Tabela 4.12).

Isso significa que foram gerados 15 experimentos sem repetições com base nas médias das 1.000 simulações e de acordo com o DIC, sendo: três para  $n = 18$ , três para  $n = 24$ , três para

$n = 30$  e seis para  $n = 36$ . Para avaliá-los, foram estabelecidas as combinações entre os níveis de dois fatores, como seguem:

c: número de combinações entre os níveis quantitativos de  $X_1$  e  $X_2$ ; e

d: DCCR, sendo 0 para não e 1 para sim.

Para cada uma das quatro medidas avaliadas ( $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$ ,  $EPMA_{\beta_2}$  e R) e dentro do valor de  $n = 36$ , foi realizada uma análise de superfície de resposta para verificar os efeitos do número de combinações entre os níveis quantitativos de  $X_1$  e  $X_2$  (c) e do DCCR (d), cujo maior modelo adotado foi definido por:

$$y_{ij} = \beta_0 + \beta_1 a_i + \beta_2 b_j + \beta_3 a_i b_j + \varepsilon'_{ij}, \text{ em que:}$$

$y_{ij}$ : valor observado da medida avaliada na combinação entre os níveis relacionados aos números de combinações entre os níveis quantitativos de  $X_1$  e  $X_2$ , isto é, aos valores de c ( $a_i$ ) [4, 6, 9, 12 e 18 ( $n = 36$ )] e aos valores do d ( $b_j$ ) (0 e 1);

$\beta_0$ : constante da regressão;

$\beta_1$ ,  $\beta_2$  e  $\beta_3$ : coeficientes da regressão; e

$$\varepsilon'_{ij} \sim N(0, \sigma_{\varepsilon'}^2).$$

Posteriormente, para o ajuste do melhor modelo, foram retirados, um de cada vez, e começando pelo mais complexo de interpretação, pelo menos um efeito não significativo, caso haja, de acordo com o teste t de Student a 5% de significância.

Já para cada uma das quatro medidas avaliadas ( $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$ ,  $EPMA_{\beta_2}$  e R) dentro dos valores de  $n = 24$  e de  $n = 30$ , separadamente, foi realizada uma análise de regressão linear para verificar somente os efeitos de c, cujo modelo adotado foi definido por:

$$y_{ij} = \beta_0 + \beta_1 a_i + \varepsilon'_{ij}, \text{ em que:}$$

$y_{ij}$ : valor observado da medida avaliada no nível relacionado ao número de níveis quantitativos ( $a_i$ ) [4, 6 e 12 ( $n = 24$ ) e 5, 6 e 15 ( $n = 30$ )];

$\beta_0$ : constante da regressão;

$\beta_1$ : coeficientes da regressão; e

$$\varepsilon'_{ij} \sim N(0, \sigma_{\varepsilon'}^2).$$

Já para  $n = 18$ , realizou-se apenas uma estatística descritiva para cada medida.

As análises estatísticas realizadas dentro de cada valor de  $n$  (18, 24, 30 e 36) tiveram como objetivo verificar se, para diferentes experimentos fatoriais instalados sob o DBC, seria melhor avaliar menos níveis quantitativos com mais repetições ou mais níveis

quantitativos com menos repetições, considerando-se o mesmo número de unidades experimentais (n) em uma análise de regressão linear múltipla.

Todas as simulações e análises estatísticas referentes às avaliações do modelo de regressão linear múltipla foram realizadas no R versão 4.0.2 (R CORE TEAM, 2020).

## 5. RESULTADOS E DISCUSSÃO

### 5.1. Regressão linear simples

Para todos os experimentos instalados sob o DBC com o mesmo número de unidades experimentais ( $n$ ), iguais a 12, 16, 20, 24, 28 e 32, as médias das medidas avaliadas  $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$  e  $EPMA_{\mu}$  diminuíram (valor  $p < 0,05$ ) em função das diminuições do número de níveis quantitativos de  $X_1$  (tratamentos) e do coeficiente de variação residual ( $CV_{\varepsilon'}$ ), em percentual (Tabela 5.1).

Conseqüentemente, para um mesmo valor de  $n$ , quanto menor for o número de níveis quantitativos ( $t = 2$ ) e o  $CV_{\varepsilon'}$  ( $cv = 10$ ), menores foram os desvios médios absolutos das estimativas de  $\beta_0$  e de  $\beta_1$  em relação aos seus respectivos parâmetros, como também, as diferenças absolutas entre os valores ajustados da variável dependente ( $Y$ ) e as respectivas médias paramétricas condicionadas a cada nível da variável independente ( $X_1$ ). Portanto, quanto menor for o número de níveis quantitativos combinado com o maior número de repetições possível, planejados em um experimento instalado sob o DBC, melhor será a qualidade de ajuste da análise de regressão linear simples. Do mesmo modo, Mateus et al. (2001), ao compararem os coeficientes de variação iguais a 3, 6, 10, 15 e 24%, concluíram que quando o coeficiente de variação experimental for maior que 6%, será necessário usar mais repetições por tratamento.

Tabela 5.1 - Superfícies de respostas ajustadas de  $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$  e  $EPMA_{\mu}$  em função do número de níveis quantitativos e do coeficiente de variação para cada valor de n

Medida	n	Superfície de resposta	R <sup>2</sup>
$EPMA_{\beta_0}$	12	$-1,8125 + 0,5449*t + 0,5496*cv$	0,99
	16	$-1,5541 + 0,3635*t + 0,4805*cv$	0,98
	20	$-1,4631 + 0,3297*t + 0,4326*cv$	0,97
	24	$-1,2641 + 0,2029*t + 0,4219*cv$	0,98
	28	$-1,1348 + 0,1740*t + 0,3822*cv$	0,98
	32	$-1,0931 + 0,1126*t + 0,3774*cv$	0,97
$EPMA_{\beta_1}$	12	$-4,7581 + 1,4045*t + 0,8458*cv$	0,97
	16	$-5,0579 + 0,9636*t + 0,8063*cv$	0,95
	20	$-3,5685 + 0,7204*t + 0,7029*cv$	0,95
	24	$-2,9740 + 0,5081*t + 0,6669*cv$	0,95
	28	$-2,9765 + 0,4401*t + 0,6234*cv$	0,93
	32	$-2,3690 + 0,3232*t + 0,5828*cv$	0,93
$EPMA_{\mu}$	12	$-0,5586 + 0,1752*t + 0,3024*cv$	0,99
	16	$-0,4605 + 0,1185*t + 0,2598*cv$	0,99
	20	$-0,3804 + 0,0960*t + 0,2328*cv$	0,99
	24	$-0,3879 + 0,0621*t + 0,2249*cv$	0,99
	28	$-0,2707 + 0,0541*t + 0,1981*cv$	0,99
	32	$-0,3986 + 0,0381*t + 0,1994*cv$	0,99

\*: significativo pelo teste t de Student (valor  $p < 0,05$ ); t = número de níveis quantitativos de  $X_1$  [ $2 \leq t \leq 6$  (n = 12),  $2 \leq t \leq 8$  (n = 16),  $2 \leq t \leq 10$  (n = 20),  $2 \leq t \leq 12$  (n = 24),  $2 \leq t \leq 14$  (n = 28) e  $2 \leq t \leq 16$  (n = 32)]; cv = coeficiente de variação residual, em percentual ( $10 \leq cv \leq 30$ ).

Reforçando a melhor qualidade de ajuste para os experimentos instalados sob o DBC com o mesmo número de unidades experimentais, quando planejados com menores valores de t e instalados sob condições com menos efeitos não controláveis, a média da medida avaliada R aumentou (valor  $p < 0,05$ ) em função das diminuições do número de níveis quantitativos e do coeficiente de variação residual, em percentual. Consequentemente, maior foi o grau de explicação de  $X_1$  sobre Y, ou seja, maior foi a proximidade do modelo ajustado em relação ao

modelo verdadeiro. Por outro lado, a média da medida avaliada ER só aumentou (valor  $p < 0,05$ ) em função da diminuição do coeficiente de variação, com exceção de  $n = 12$  (Tabela 5.2).

Tabela 5.2 - Superfícies de respostas ajustadas de R e ER em função do número de níveis quantitativos e do coeficiente de variação para cada valor de n

Medida	n	Superfície de resposta	R <sup>2</sup>
R	12	$1,2702 - 0,0495*t - 0,0175*cv$	0,94
	16	$1,2477 - 0,0363*t - 0,0178*cv$	0,91
	20	$1,2283 - 0,0280*t - 0,0192*cv$	0,90
	24	$1,2177 - 0,0237*t - 0,0197*cv$	0,91
	28	$1,2027 - 0,0193*t - 0,0197*cv$	0,88
	32	$1,1976 - 0,0166*t - 0,0198*cv$	0,87
ER	12	$13,2530 - 0,2999*t - 0,3665*cv$	0,85
	16	$7,1687 - 0,2018*cv$	0,83
	20	$4,7880 - 0,1231*cv$	0,86
	24	$3,9717 - 0,0976*cv$	0,86
	28	$3,4148 - 0,0794*cv$	0,86
	32	$2,9808 - 0,0647*cv$	0,86

\*: significativo pelo teste t de Student (valor  $p < 0,05$ ); t = número de níveis quantitativos de  $X_1$  [ $2 \leq t \leq 6$  ( $n = 12$ ),  $2 \leq t \leq 8$  ( $n = 16$ ),  $2 \leq t \leq 10$  ( $n = 20$ ),  $2 \leq t \leq 12$  ( $n = 24$ ),  $2 \leq t \leq 14$  ( $n = 28$ ) e  $2 \leq t \leq 16$  ( $n = 32$ )]; cv = coeficiente de variação residual, em percentual ( $10 \leq cv \leq 30$ ).

Quanto maior for a ER, mais eficiente será o DBC em relação ao DIC. De acordo com os resultados, essa eficiência aumentou quando os diferentes blocos com t unidades experimentais homogêneas foram avaliadas sob as ocorrências de menores efeitos dos fatores não controláveis, isto é, nos experimentos com os menores coeficientes de variação residuais. Concluiu-se que a maior redução relativa da variância residual (QMResReg), principal vantagem de se utilizar um DBC ao invés de um DIC (SHIEH; JAN, 2004), ocorreu na presença de menores CVs. Além disso, notou-se, também, que a ER não se relacionou, nem com o número e nem com o tamanho dos blocos.

Desse modo, concluiu-se que, entre os experimentos avaliados, 10 ( $n = 20$ ),  $t = 2$  e  $r = 12$  ( $n = 24$ ),  $t = 2$  e  $r = 14$  ( $n = 28$ ) e  $t = 2$  e  $r = 16$  ( $n = 32$ ), melhor foi o que apresentou desempenho da análise de regressão linear simples. Conseqüentemente, caso haja alguma expectativa do ajuste deste modelo, recomenda-se experimentar apenas os níveis referentes aos limites inferior e superior do intervalo da variável independente  $X_1$ . Isso significa que quanto mais repetições (blocos) de um mesmo nível quantitativo de  $X_1$  forem feitas, melhor. Storck et al. (2009), avaliaram a produtividade de grãos de soja em 216 ensaios de competição de genótipos executados sob o DBC e recomendaram quatro repetições por tratamento conforme o método de Papadakis e 11 sem a utilização deste método.

Nesse sentido, não há necessidade de serem avaliados mais níveis quantitativos de  $X_1$ , com exceção dos limites inferior (LI) e superior (LS), a fim de proporcionar menos distâncias entre os níveis intermediários e localizados dentro do intervalo avaliado da variável independente. Na verdade, não há necessidade de serem avaliados quaisquer um dos seguintes níveis quantitativos de  $X_1$ :

$$LI < x_{1_i} < LS.$$

Por outro lado, caso não haja, *a priori*, uma expectativa do ajuste do modelo linear, recomendam-se três tratamentos quantitativos e não mais do que isso, sendo eles definidos por:

$$x_{1_1} = LI;$$

$$x_{1_2} = PC; \text{ e}$$

$$x_{1_3} = LS, \text{ em que:}$$

PC: ponto central (média dos limites inferior e superior).

Na relação linear do Y em função de  $X_1$ , existem os modelos verdadeiro ( $\mu_i = \beta_0 + \beta_1 x_{1_i}$ ) e ajustado ( $\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1_i}$ ) que, na grande maioria das vezes, são diferentes. Portanto, como para o ajuste de uma reta, há necessidade de apenas dois pontos (níveis quantitativos de  $X_1$ ), os resultados mostraram que foi preferível estimar as médias desses pontos com mais repetições. Caso contrário, quantos mais pontos, ou seja, quanto mais níveis quantitativos de  $X_1$  com as suas respectivas médias estimadas, conseqüentemente, com menos repetições, pior foi o ajuste linear. Isso significa que quanto mais repetições estiver associada à estimativa de uma média, menos erro associado a ela ocorrerá. Para ilustrar essa conclusão, foram construídos os gráficos de Y em função de  $X_1$ , de acordo com a primeira das 1.000 simulações realizadas com  $n = 32$  e  $cv = 30\%$ , para  $t = 2$  e  $r = 16$  (Figura 5.1a) e para  $t = 16$  e  $r = 2$  (Figura 5.1b).

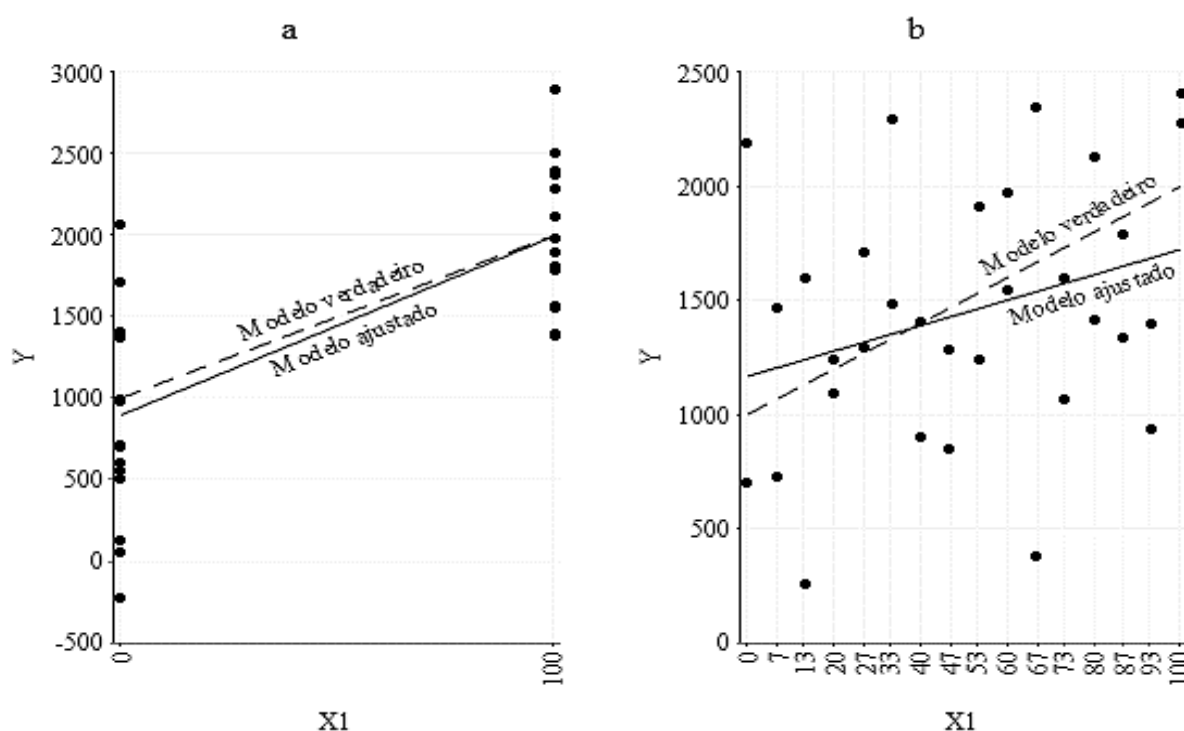


Figura 5.1 - Modelos lineares verdadeiro ( $\mu_i = 1.000 + 10x_{1i}$ ) e ajustado de Y em função de  $X_1$  para  $t = 2$  e  $r = 16$  ( $\hat{y}_{ij} = 896,5323 + 10,9832x_{1i}$ ) (a) e para  $t = 16$  e  $r = 2$  ( $\hat{y}_{ij} = 1.167,1042 + 5,5717x_{1i}$ ) (b).

Além disso, como já era esperado, quanto menor foi o  $CV_{\epsilon'}$ , melhor foi ajuste da equação de regressão e mais precisas foram as estimativas dos seus parâmetros. Quanto menor for o  $CV_{\epsilon'}$ , menor será a variação residual (QMResReg) e, portanto, maior será a precisão experimental. Consequentemente, as médias de tratamentos terão estimativas com menores erros, mesmo para aquelas estimadas com os menores números de repetições.

O  $CV_{\epsilon'}$ , quando estimado, expressa a raiz do QMResReg em percentual da estimativa da média geral de Y obtida no experimento. Por ser adimensional, ele pode ser utilizado para dar uma ideia da precisão do experimento. Para os coeficientes de variação obtidos mais comumente nos experimentos de campo das Ciências Agrárias, podem-se considerá-los baixos, médios, altos e muito altos, quando inferiores a 10%, situados entre 10% e 20%, situados entre 20% e 30% e quando superiores a 30%, respectivamente. Consequentemente, podem-se caracterizar os respectivos experimentos como de precisões alta, média, baixa e muito baixa (GOMES, 2009).

Em igualdade de condições, será mais preciso o experimento que proporcionar menor coeficiente de variação. Porém, o experimento com maior número de repetições por tratamento também terá expectativa de ser mais preciso. Assim, para ser obtida uma mesma precisão em relação à estimativa de  $\beta_0$  para experimentos instalados sob o DBC com  $n = 32$ , pode-se utilizar, por exemplo, as seguintes combinações:  $t = 2$  ( $r = 16$ ) e  $cv = 14,18$ ,  $t = 8$  ( $r = 4$ ) e  $cv = 12,39$  e  $t = 16$  ( $r = 2$ ) e  $cv = 10$  (Tabela 5.1). E para à de  $\beta_1$ , são mencionadas as seguintes:  $t = 2$  ( $r = 16$ ) e  $cv = 17,77$ ,  $t = 8$  ( $r = 4$ ) e  $cv = 14,44$  e  $t = 16$  ( $r = 2$ ) e  $cv = 10$  (Tabela 5.1).

## 5.2. Regressão linear múltipla

Para  $n = 18$ , as maiores estimativas de  $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$  e  $EPMA_{\beta_2}$  foram proporcionadas pelo DCCR, caracterizando-o como, provavelmente, menos eficiente em prover estimativas mais próximas dos parâmetros  $\beta_0$ ,  $\beta_1$  e  $\beta_2$ , quando comparado aos delineamentos  $2 \times 3$  e  $3 \times 3$  (DCCFC). E em relação aos dois últimos, o fatorial  $2 \times 3$  se mostrou, ligeiramente, melhor (Figura 5.2).

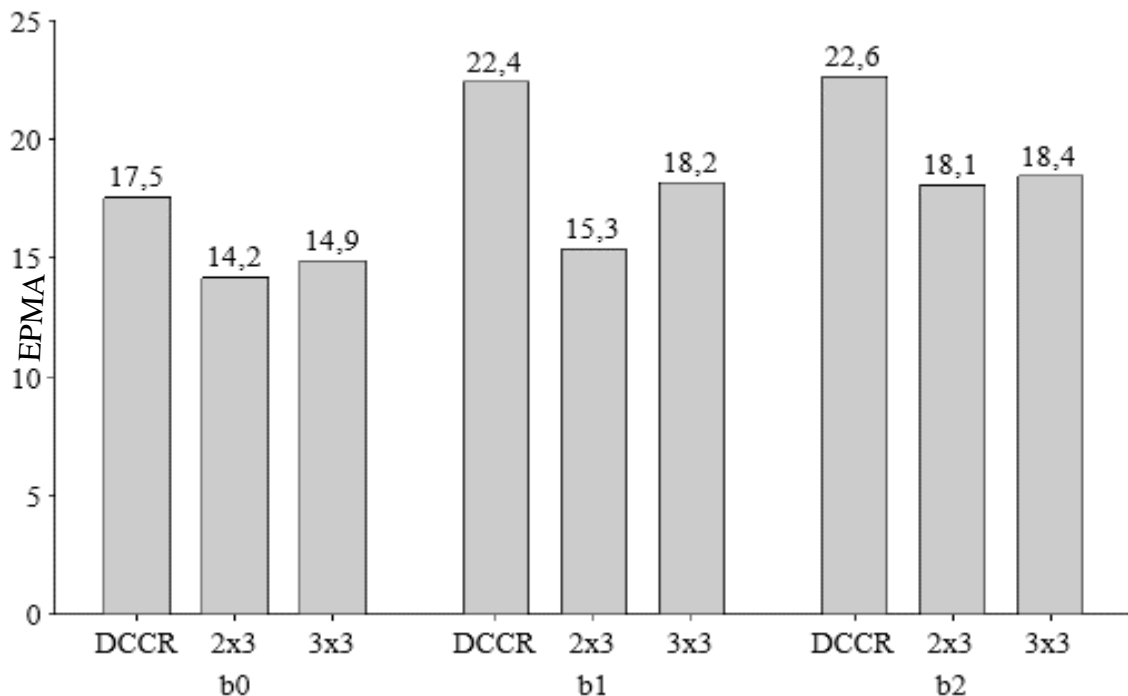


Figura 5.2 - Estimativas de  $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$  e  $EPMA_{\beta_2}$  proporcionadas pelo DCCR,  $2 \times 3$  e  $3 \times 3$  (DCCFC), para  $n = 18$ .

Reforçando o melhor desempenho do fatorial  $2 \times 3$ , tem-se a maior estimativa da média da medida avaliada  $R$ , que está relacionada ao grau de explicação da equação de regressão ajustada sobre a variável dependente  $Y$  (Figura 5.3). Isso confirma que quanto menor for o número de níveis quantitativos combinado com o maior número de repetições possível, quando planejados em um experimento instalado sob o DBC, melhor será a qualidade de ajuste da análise de regressão linear múltipla.

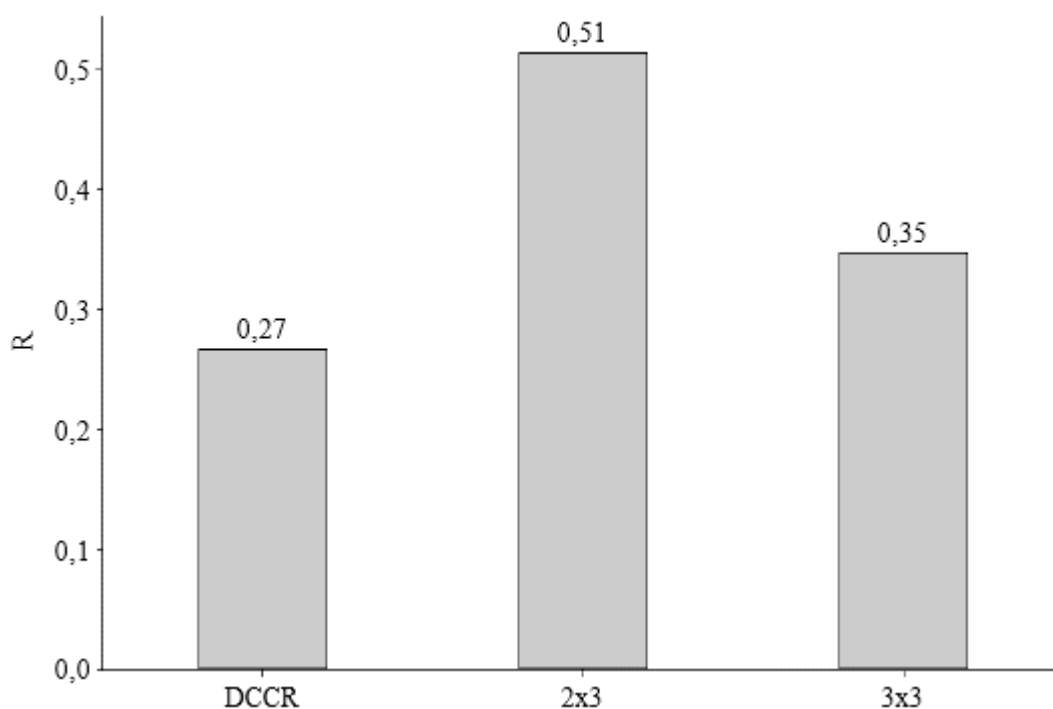


Figura 5.3 - Estimativas de  $R$  proporcionadas pelo DCCR,  $2 \times 3$  e  $3 \times 3$  (DCCFC), para  $n = 18$ .

Nesse caso, deve-se recomendar o fatorial  $2 \times 3$ , ao invés do DCCR e do  $3 \times 3$  (DCCFC). Do mesmo modo, Penteadó e Batista (1971) obtiveram resultados similares aos encontrados, quando compararam as eficiências do DCCR e do fatorial  $5 \times 5$ . Além deles, Campos (1967), ao comparar o fatorial  $3 \times 3$  (DCCFC) e o DCCR, concluiu que o primeiro foi mais preciso, baseando-se nas estimativas das variâncias dos estimadores da superfície de resposta.

Já para os valores de  $n$  iguais a 24 e 30 ( $EPMA_{\beta_0}$ ) e, para  $n = 30$  ( $EPMA_{\beta_2}$ ), as médias das duas medidas avaliadas diminuíram (valor  $p < 0,05$ ) em função das diminuições do número de combinações entre os níveis quantitativos de  $X_1$  e  $X_2$  (tratamentos). E para  $EPMA_{\beta_1}$  e  $EPMA_{\beta_2}$  ( $n = 24$ ), as médias estimadas não sofreram (valor  $p > 0,05$ ) alterações. Por outro lado, a média da medida avaliada  $R$  aumentou (valor  $p < 0,05$ ) em função das diminuições do número de tratamentos, tanto para  $n = 24$  como para  $n = 30$  (Tabela 5.3).

Tabela 5.3 - Equações de regressão ajustadas de  $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$ ,  $EPMA_{\beta_2}$  e R em função do número de combinações entre os níveis quantitativos para cada valor de n

Medida	n	Equação de regressão	R <sup>2</sup>
$EPMA_{\beta_0}$	24	$10,6065 + 0,2197*c$	0,99
	30	$-11,0025 + 4,0688*c$	0,99
$EPMA_{\beta_1}$	24	14,05	–
	30	13,12	–
$EPMA_{\beta_2}$	24	15,39	–
	30	$12,1150 + 0,2617*c$	1,00
R	24	$0,8475 - 0,0428*c$	0,99
	30	$0,7966 - 0,0311*c$	0,99

\*: significativo pelo teste t de Student ( $P < 0,05$ ); c = número de combinações entre os níveis quantitativos de  $X_1$  e  $X_2$  [ $4 \leq c \leq 12$  (n = 24) e  $5 \leq c \leq 15$  (n = 30)].

Para todos os experimentos instalados sob o DBC com o mesmo número de unidades experimentais (n) igual a 36, as médias das medidas avaliadas  $EPMA_{\beta_0}$  e  $EPMA_{\beta_2}$  diminuiram (valor  $p < 0,05$ ) em função das diminuições do número de combinações entre os níveis quantitativos de  $X_1$  e  $X_2$  e da ausência do DCCR. Já para  $EPMA_{\beta_1}$ , a média diminuiu (valor  $p < 0,05$ ) apenas em função da ausência do DCCR. Por outro lado, a média da medida avaliada R aumentou (valor  $p < 0,05$ ) em função das diminuições do número de tratamentos e da presença do DCCR (Tabela 5.4).

Conseqüentemente, para um mesmo valor de n, quanto menor forem o número de combinações entre os níveis quantitativos e sem a utilização do DCCR, menores serão as estimativas dos desvios médios absolutos em relação aos respectivos parâmetros  $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$  e  $EPMA_{\beta_2}$ , como também, as diferenças absolutas entre os valores ajustados e verdadeiros da variável dependente (Y). Portanto, quanto menor for o número de níveis quantitativos combinado com o maior número de repetições possível, quando planejados em um experimento instalado sob o DBC, melhor será a qualidade de ajuste da análise de regressão linear múltipla.

Tabela 5.4 - Superfícies de resposta ajustadas de  $EPMA_{\beta_0}$ ,  $EPMA_{\beta_1}$ ,  $EPMA_{\beta_2}$  e R em função do número de combinações entre os níveis quantitativos e do DCCR para  $n = 36$

	Superfície de resposta	$R^2$
$EPMA_{\beta_0}$	$8,9625 + 0,1342*c + 2,1325*d$	0,98
$EPMA_{\beta_1}$	$11,9548 + 3,9573*d$	0,67
$EPMA_{\beta_2}$	$9,9135 + 0,3130*c + 3,0606*d$	0,95
R	$0,8093 - 0,0278*c - 0,1226*d$	0,98

\*: significativo pelo teste t de Student (valor  $p < 0,05$ );  $c$  = número de combinações entre os níveis quantitativos de  $X_1$  e  $X_2$  ( $4 \leq c \leq 18$ );  $d$  = delineamento composto central rotacional (0 = não e 1 = sim).

Desse modo, concluiu-se que, entre os experimentos avaliados, quanto menor foi o número de combinações entre os níveis quantitativos e maior foi o número de blocos para um mesmo número de unidades experimentais [ $c = 4$  ( $2 \times 2$ ) e  $r = 6$ ,  $c = 5$  ( $2^2 + 1$ ) e  $r = 6$ ,  $c = 4$  ( $2 \times 2$ ) e  $r = 8$ ], melhor foi o desempenho da análise de regressão linear múltipla. Consequentemente, caso haja alguma expectativa do ajuste deste modelo, recomenda-se experimentar apenas os níveis referentes aos limites inferior e superior do intervalo das variáveis independentes  $X_1$  e  $X_2$ . Isso significa que quanto mais repetições (blocos) de um mesmo nível quantitativo de  $X_1$  e de um mesmo nível quantitativo de  $X_2$  forem feitas, melhor.

Nesse sentido, não há necessidade de serem avaliados mais níveis quantitativos de  $X_1$  e de  $X_2$ , com exceção dos seus respectivos limites inferior (LI) e superior (LS), a fim de proporcionar menos distâncias entre os níveis intermediários e localizados dentro do intervalo avaliado das variáveis independentes. Na verdade, não há necessidade de serem avaliados quaisquer um dos seguintes níveis quantitativos de  $X_1$  e  $X_2$ :

$$LI_1 < x_{1_i} < LS_1; \text{ e}$$

$$LI_2 < x_{2_j} < LS_2.$$

Por outro lado, caso não haja, *a priori*, uma expectativa do ajuste do modelo linear, recomendam-se nove tratamentos quantitativos oriundos de um fatorial  $3 \times 3$  (DCCFC) e não mais do que isso, sendo os níveis de  $X_1$  e  $X_2$  definidos por:

$$x_{1_1} = LI_1; x_{1_2} = PC_1; x_{1_3} = LS_1; \text{ e}$$

$$x_{2_1} = LI_2; x_{2_2} = PC_2; x_{2_3} = LS_2, \text{ em que:}$$

$PC_1$  e  $PC_2$ : pontos centrais de  $X_1$  (média de  $LI_1$  e  $LS_1$ ) e  $X_2$  (média de  $LI_2$  e  $LS_2$ ).

## 6. CONCLUSÕES

Para o ajuste de um modelo de regressão linear simples em um experimento instalado sob o DBC, a qualidade aumenta com a diminuição do número de níveis quantitativos e com o aumento do número de repetições por nível quantitativo. Isso implica em recomendar, para um mesmo número de unidades experimentais, o menor número possível de níveis quantitativos. Se houver uma expectativa para o modelo linear, pode-se então, recomendar apenas dois níveis quantitativos. Caso contrário, recomendam-se três.

Para o ajuste de um modelo de regressão linear múltipla com duas variáveis independentes em um experimento instalado sob o DBC, a qualidade aumenta com a diminuição do número de combinações entre os níveis quantitativos e com o aumento do número de repetições por combinação. Isso implica em recomendar, para um mesmo número de unidades experimentais, o menor número possível de combinações entre os níveis quantitativos das duas variáveis independentes. Se houver uma expectativa para o modelo com apenas efeitos lineares, pode-se então, recomendar apenas dois níveis quantitativos por variável independente avaliados em um fatorial  $2 \times 2$ . Caso contrário, recomendam-se três, sendo avaliados em um fatorial  $3 \times 3$ .

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola**. 4. ed. Jaboticabal: FUNEP, 2006. 237 p.
- BAHRY, C. A.; VENSKE, E.; NARDINO, M.; FIN, S. S.; ZIMMER, P. D.; SOUZA, V. Q.; CARON, B. O. Características morfológicas e componentes de rendimento da soja submetida à adubação nitrogenada. **Revista Agrarian**, v. 6, p. 281-288, 2013.
- BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. **Como fazer experimentos – pesquisa e desenvolvimento na ciência e na indústria**. 2. ed. Campinas: Editora UNICAMP, 2003. 401 p.
- BASTOS, E. V. P.; GUIMARÃES, J. C. F.; SEVERO, E. A. **Modelo de regressão linear para análise de investimentos em uma empresa do ramo petrolífero**. Rio de Janeiro: Revista Produção e Desenvolvimento, 2015. 88 p.
- BONILLA, J. A. **Métodos quantitativos**. Belo Horizonte: Editora Littera Maciel, 1995. 249 p.
- BOX, G. E. P.; DRAPER, N. R. **Empirical model – building and response surfaces**. New York: John Wiley & Sons, 1987. 669 p.
- BOX, G. E. P.; WILSON, K. B. On the experimental attainment of optimum conditions. **Journal of the Royal Statistical Society**, 1951. 45 p.
- CAMPOS, H. **Aspectos da aplicação das superfícies de resposta a ensaios fatoriais 3<sup>3</sup> de adubação**. Piracicaba: teste de doutorado – ESALQ/USP, 1967. 82 p.
- COCHRAN, W. G.; COX, G. M. **Experimental designs**. 2. ed. New York: John Wiley & Sons, 1992. 611 p.
- DEVORE, J. L. **Probabilidade e estatística para engenharia e ciências**. 6. ed. São Paulo: Cengage Learning, 2012. 692 p.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: John Wiley & Sons, 1998. 693 p.
- GEA, M. M.; BATANERO, C.; CANADAS, G.; CONTRERAS, J. M. Um estudio empírico de las situaciones-problema de correlación y regresión em libros de texto de Bachillerato. **Investigación en Educación Matemática**, 2013. 300 p.
- GOMES, F. P. **Curso de estatística experimental**. 15. ed. São Paulo: Livraria Universo Agrícola, 2009. 451 p.
- GUPTA, B. C.; GUTTMAN, I. **Estatística e probabilidade com aplicações para engenheiros e cientistas**. Rio de Janeiro: LTC, 2017. 748 p.

- HAALAND, P. D. **Experimental design in biotechnology**. New York: Marcel Dekker Inc, 1989. 259 p.
- HOFFMANN, R.; VIEIRA, S. **Análise de regressão: uma introdução à econometria**. 2. ed. São Paulo: Hucitec, 1983. 379 p.
- KRAJEWSKI, L. J.; RITZMAN, L. P.; MALHOTRA, M. **Administração da produção e operações**. 8. ed. São Paulo: Pearson Prentice Hall, 2009. 640 p.
- LIMA, P. C.; ABREU, A. R. **Delineamento e análise de experimentos**. Lavras: FAEPE, 2000. 45 p.
- MAROCO, J. **Análise estatística com utilização do SPSS**. 3. ed. Lisboa: Edições Sílabo, 2010. 824 p.
- MASON, R. L.; GUNST, R. L.; HESS, J. L. **Statistical design and analysis of experiments with applications to engineering and Science**. 2. ed. Hoboken: John Wiley & Sons, 2003. 760 p.
- MATEUS, N. B.; BARBIN D.; CONAGIN, A. O delineamento composto central e sua viabilidade de uso em algumas áreas de pesquisa. **Revista Acta Scientiarum**, v. 23, p. 1537-1546, 2001.
- MONTGOMERY, D. C. **Design and analysis of experiments**. 7. ed. New York: John Wiley & Sons, 2009. 656 p.
- MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 5. ed. Rio de Janeiro: LTC, 2012. 521 p.
- MYERS, R. H.; MONTGOMERY, D. C.; ANDERSON-COOK, C. M. **Response surface methodology: process and product optimization using designed experiments**. 3. ed. New Jersey: John Wiley & Sons, 2009. 680 p.
- OLIVEIRA, M. W.; TRIVELIN, P. C. O.; GAVA, G. J. C.; PENATTI, C. P. **Degradação da palhada da cana-de açúcar**. Piracicaba: Scientia Agrícola, 1999. 809 p.
- PENTEADO, A. F.; BATISTA, L. B. Eficiência do ensaio composto central (Box) em comparação com os fatoriais completos de dois fatores. **Congresso Brasileiro de Ciência do Solo**, v. 13, 1971.
- POSSATO JUNIOR, O.; BERTAGNA, F. A. B.; PETERLINI, E.; BALERONI, A. G.; ROSSI, R. M.; ZENI NETO, H. Survey of statistical methods applied in articles published in Acta Scientiarum. **Acta Scientiarum. Agronomy**, 2019. 10 p.
- R Core Team. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2020. URL <https://www.r-project.org>.

- RIBEIRO JÚNIOR, J. I. **Métodos estatísticos aplicados à melhoria da qualidade**. Viçosa: Editora UFV, 2012. 385 p.
- RODRIGUES, M. I.; IEMMA, A. F. **Planejamento de experimentos e otimização de processos**. Campinas: Editora UNICAMP, 2005. 326 p.
- SHIEH, G.; JAN, S. L. **The effectiveness of randomized complete block design**. Neerlandica: Statistica Neerlandica, 2004. 124 p.
- STORCK, L. **Experimentação vegetal**. 2. ed. Santa Maria: UFSM, 2006. 198 p.
- WERKEMA, M. C. C.; AGUIAR, S. **Análise de regressão: como entender o relacionamento entre as variáveis de um processo**. Belo Horizonte: Werkema Editora, 2006. 306 p.
- YATES, F. **The design and analysis of factorial experiments**. Harpend: Imperial Bureau of Soil Science, 1937. 96 p.
- ZEVIANE, W. M. **Manual de planejamento e análise de experimentos com R**. Curitiba: UFPR, 2011. 276 p.
- ZIMMERMANN, F. J. P. **Estatística aplicada à pesquisa agrícola**. 2. ed. Brasília: EMBRAPA, 2014. 582 p.