PERFORMANCE OF MULTIPLE LINEAR REGRESSION ANALYSIS CONDUCTED UNDER RANDOMIZED COMPLETE BLOCK DESIGN

Daibou Alassane[1], Jaqueline Akemi Suzuki Sediyama[1], Alice dos Santos Ribeiro[1], José Ivo Ribeiro Júnior[1], Belo Afonso Muetanene[2]

[1] *UFV – Universidade Federal de Viçosa, Minas Gerais State, Brazil. E-mails: daibou.alassane@ufv.br, alice.ribeiro@ufv.br, jivo@ufv.br, jaqueline.suzuki@ufv.br*
[2] *Rural Engineering Department, Faculty of Agronomic Sciences, Lúrio University, Mozambique, E-mail: floriafonso@gmail.com*

ABSTRACT

In factorial experiments conducted under randomized block design, the multiple linear regression model fitting can be performed under different combinations of the quantitative levels of the two factors and the number of replications. To determine the best combination, considering the same number of levels per factor and the same number of experimental units, it was concluded through a simulated data study that the quality of the fit increases when regression is performed in experiments with fewer combinations of levels (treatments) and more replications. Therefore, if linearity is expected, using four treatments evaluated in a 2 × 2 factorial design for model fitting is recommended. Otherwise, nine treatments evaluated in a 3 × 3 factorial design are recommended. All of this is for experiments with coefficients of variation of 20%.

Keywords: Treatments, replications, experimental precision.

INTRODUCTION

In agricultural sciences, many double factorial experiments are conducted with quantitative levels (treatments) under randomized complete block design (RCBD), where the responses are analyzed using the multiple linear regression model. And, To fit this model, one should, a priori, determine the dependent variable (Y) and the range of values for the two independent variables (X_1 and X_2) defined by their respective lower and upper limits (MONTGOMERY, 2009, 2012; POSSATO et al., 2019).

In addition to the full factorial design used to obtain an experiment with two independent variables, X_1 and X_2, incomplete factorials can also be used, which save the number of combinations (treatments). The central composite design (CCD) is a design composed of cube points (±1), axial points (±α), and central points (0) in a highly flexible manner. It is a simple and economical design, widely used in response surface methodology. According to Myers et al. (2009), the most common value of α is the one that makes the CCD rotatable (RCCD). When the RCCD is complete for two independent variables (X_1 and X_2), we have α = 1.4142. And when the chosen value is α = 1, the axial points will be located on the cube, and the design will be referred to as the face-centered central composite design (FCCD).

As previously mentioned, the quantitative levels of the two factors should include the lower and upper limits and at least one intermediate level, as the objective is to fit a multiple linear regression model that encompasses the entire studied range. This allows for great flexibility in

DOI: 10.37856/bja.v98i3.4334

choosing the number of treatments. Therefore, the experiments are divided into two main classes for the same number of experimental units: a greater number of treatments with fewer replications per treatment and a smaller number of treatments with more replications per treatment.

Thus, more cost-effective double factorial experiments with less than four replications per treatment (combinations of quantitative levels of the two factors) planned under randomized block design and analyzed using multiple linear regression can also be good options. Therefore, this study aimed to evaluate the effects of the number of treatments and replications on the performance of the linear regression model with two independent variables using simulated data from experiments conducted under a randomized block design.

MATERIAL AND METHODS

Regression Parameters

The multiple linear regression model that represented the functional relationship between the dependent variable (Y) and the two independent (X_1 e X_2) was given by:

y_{ijk} = 1.000 + $10x_{1i}$ + $10x_{2j}$ + ε_{ijk}, for $0 \leq X_1 \leq 100$ e $0 \leq X_2 \leq 100$, where:

y_{ijk}: the observed value of the dependent variable Y at the combination of quantitative levels x_{1i} (i = 1, 2, ..., t) and x_{2j} (j = 1, 2, ..., g) and in block b_k (k = 1, 2, ..., r);

β_0 = 1.000: regression constant;

$\beta_1 = \beta_2$ = 10: regression coefficients.;

ε_{ijk}: regression error associated with the observed value y_{ijk};

μ_{ij} = 1.000 + $10x_{1i}$ + $10x_{2j}$: population mean of the dependent variable Y at the combination of quantitative levels x_{1i} e x_{2j}; and

μ = 2.000: overall population means of the dependent variable Y.

Data Simulation

To obtain the regression residuals (e_{ijk}), 1,000 simulations were performed according to the normal distribution with a population mean of zero and a population standard deviation σ_ε, where:

e_{ijk}: regression residual associated with the observed value y_{ijk} (i = 1, 2, ..., and j = 1, 2, ..., and k = 1, 2, ..., r).

Initially, the value was defined for the simulation realizations to provide residual coefficients of variation (CV_ε) equal to 20%, according to the following expression.:

$$CV_\varepsilon = 100 \times \frac{\sigma_\varepsilon}{\mu} = 100 \times \frac{\sigma_\varepsilon}{2000}.$$

Therefore, for the 1,000 simulation realizations, a value equal to 400 was adopted. Consequently, the following normal distribution is obtained:

$$\varepsilon_{ijk} \sim N\,(\mu_\varepsilon = 0;\ \sigma_\varepsilon^2 = 400^2).$$

Randomized Complete Block Design

For comparison purposes, 15 double factorial experiments were conducted under a randomized complete block design (RCBD), considering 15 combinations between the number of treatments (combinations between quantitative levels of X_1 and X_2 ranging from zero to 100) (n_{trat}) and the number of blocks (r), to provide the same numbers of experimental units (n) equal to 18, 24, 30, and 36.

In this study, three experiments were conducted under the randomized complete block design (RCBD) for n = 18 (Table 1), three for n = 24 (Table 2), three for n = 30 (Table 3), and six for n = 36 (Table 4), as follows:

x_{1i}: quantitative level of the independent variable X_1 (i = 1, 2, ...,);

x_{2j}: quantitative level of the independent variable X_2 (j = 1, 2, ...,);

ω_k: effect of block b_k (k = 1, 2, ..., r);

The block effects, taking the experiments installed under RCBD with CV_ε = 20% as references, were defined to provide approximately the same block sum of squares (SSBl) that would promote significance for the block effects themselves ($f_{calBl} \geq f_{tabBl}$) in all experiments with the same value of n and α = 0.05.

n_{trat}=6 (2x3) e r=3		n_{trat}=9 (RCCD) and r=2		n_{trat}=9 (FCCD) and r=2	
x_{1i} | x_{2j}	ω_k	x_{1i} | x_{2j}	ω_k	x_{1i} | x_{2j}	ω_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	–

Table 1. Combinations of quantitative levels and block effects for the three factorial experiments conducted under RCBD with n = 18.

n_{trat}=4 (2x2) and r = 6		n_{trat}=6 (2x3) and r = 4		n_{trat}=12 (3x4) and r = 2	
x_{1i} | x_{2j}	ω_k	x_{1i} | x_{2j}	ω_k	x_{1i} | x_{2j}	ω_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	–

Table 2. Combinations of quantitative levels and block effects for the three factorial experiments conducted under RCBD with n = 24.

DOI: 10.37856/bja.v98i3.4334

Table 3. Combinations of quantitative levels and block effects for the three factorial experiments conducted under RCBD with n = 30.

<table>
<tr><th colspan="2">n_{trat}=5 (2^2+1) and r = 6</th><th colspan="2">n_{trat}=6 (2x3) and r = 5</th><th colspan="2">n_{trat}=15 (3x5) and r = 2</th></tr>
<tr><th>x_{1i} | x_{2j}</th><th>ω_k</th><th>x_{1i} | x_{2j}</th><th>ω_k</th><th>x_{1i} | x_{2j}</th><th>ω_k</th></tr>
<tr><td>0 | 0</td><td>-450</td><td>0 | 0</td><td>-500</td><td>0 | 0</td><td>-685</td></tr>
<tr><td>100 | 0</td><td>-250</td><td>100 | 0</td><td>-200</td><td>50 | 0</td><td>685</td></tr>
<tr><td>0 | 100</td><td>-100</td><td>0 | 50</td><td>0</td><td>100 | 0</td><td>–</td></tr>
<tr><td>100 | 100</td><td>100</td><td>100 | 50</td><td>200</td><td>0 | 25</td><td>–</td></tr>
<tr><td>50 | 50</td><td>250</td><td>0 | 100</td><td>500</td><td>50 | 25</td><td>–</td></tr>
<tr><td>–</td><td>450</td><td>100 | 100</td><td>–</td><td>100 | 25</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>0 | 50</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>50 | 50</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>100 | 50</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>0 | 75</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>50 | 75</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>100 | 75</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>0 | 100</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>50 | 100</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>100 | 100</td><td>–</td></tr>
</table>

Table 4. Combinations of quantitative levels and block effects for the six factorial experiments conducted under RCBD with n = 36.

<table>
<tr><th colspan="2">n_{trat}=4 (2x2) and r = 9</th><th colspan="2">n_{trat}=6 (2x3) and r = 6</th><th colspan="2">n_{trat}=9 (DCCR) and r = 4</th><th colspan="2">n_{trat}=9 (DCCFC) and r = 4</th><th colspan="2">n_{trat}=12 (3x4) and r = 3</th><th colspan="2">n_{trat}=18 (3x6) and r = 2</th></tr>
<tr><th>x_{1i} | x_{2j}</th><th>ω_k</th><th>x_{1i} | x_{2j}</th><th>ω_k</th><th>x_{1i} | x_{2j}</th><th>ω_k</th><th>x_{1i} | x_{2j}</th><th>ω_k</th><th>x_{1i} | x_{2j}</th><th>ω_k</th><th>x_{1i} | x_{2j}</th><th>ω_k</th></tr>
<tr><td>0 | 0</td><td>-380</td><td>0 | 0</td><td>-365</td><td>14,64 | 14,64</td><td>-500</td><td>0 | 0</td><td>-500</td><td>0 | 0</td><td>-590</td><td>0 | 0</td><td>-670</td></tr>
<tr><td>100 | 0</td><td>-250</td><td>100 | 0</td><td>-300</td><td>85,36 | 14,64</td><td>-235</td><td>100 | 0</td><td>-235</td><td>50 | 0</td><td>0</td><td>50 | 0</td><td>670</td></tr>
<tr><td>0 | 100</td><td>-150</td><td>0 | 50</td><td>-200</td><td>14,64 | 85,36</td><td>235</td><td>0 | 100</td><td>235</td><td>100 | 0</td><td>590</td><td>100 | 0</td><td>–</td></tr>
<tr><td>100 | 100</td><td>-100</td><td>100 | 50</td><td>200</td><td>85,36 | 85,36</td><td>500</td><td>100 | 100</td><td>500</td><td>0 | 33,33</td><td>–</td><td>0 | 20</td><td>–</td></tr>
<tr><td>–</td><td>0</td><td>0 | 100</td><td>300</td><td>0 | 50</td><td>–</td><td>0 | 50</td><td>–</td><td>50 | 33,33</td><td>–</td><td>50 | 20</td><td>–</td></tr>
<tr><td>–</td><td>100</td><td>100 | 100</td><td>365</td><td>100 | 50</td><td>–</td><td>100 | 50</td><td>–</td><td>100 | 33,33</td><td>–</td><td>100 | 20</td><td>–</td></tr>
<tr><td>–</td><td>150</td><td>–</td><td>–</td><td>50 | 0</td><td>–</td><td>50 | 0</td><td>–</td><td>0 | 66,67</td><td>–</td><td>0 | 40</td><td>–</td></tr>
<tr><td>–</td><td>250</td><td>–</td><td>–</td><td>50 | 100</td><td>–</td><td>50 | 100</td><td>–</td><td>50 | 66,67</td><td>–</td><td>50 | 40</td><td>–</td></tr>
<tr><td>–</td><td>380</td><td>–</td><td>–</td><td>50 | 50</td><td>–</td><td>50 | 50</td><td>–</td><td>100 | 66,67</td><td>–</td><td>100 | 40</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>0 | 100</td><td>–</td><td>0 | 60</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>50 | 100</td><td>–</td><td>50 | 60</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>100 | 100</td><td>–</td><td>100 | 60</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>0 | 80</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>50 | 80</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>100 | 80</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>0 | 100</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>50 | 100</td><td>–</td></tr>
<tr><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>–</td><td>100 | 100</td><td>–</td></tr>
</table>

As can be observed for n = 18 (Table 1) and n = 36 (Table 4), the FCCD is a 3 × 3 factorial design. Therefore, from now on, it has been given the final designation.

Thus, for each of the four values of n (18, 24, 30, and 36), 1,000 simulations were performed according to their respective normal distribution ($\mu_\varepsilon = 0$ e $\sigma_\varepsilon = 400$) to generate the n and their respective regression residuals.

Subsequently, the observed values of the dependent variable Y in each of the 15 balanced experiments installed under the randomized complete block design (RCBD) were obtained by:

$y_{ijk} = 1.000 + 10x_{1i} + 10x_{2j} + \omega_k + e_{ijk}$, where:

y_{ijk}: the observed value of the dependent variable Y at the combination of quantitative levels x_{1i} (i = 1, 2, ...) and x_{2j} (j = 1, 2, ...) and in block b_k (k = 1, 2, ..., r).

In total, 15 different datasets were generated for the study of multiple linear regression analysis, and for each of them, 1,000 simulations were performed.

For each of the 15,000 datasets, a multiple linear regression model was fitted as follows:

$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2j}$, where:

$\hat{y}_{ij}$: adjusted value of the dependente variable Y at the combination of quantitative levels x_{1i} (i = 1, 2, ...,) and x_{2j} (j = 1, 2, ...,).

Afterward, the regression analysis with the test for lack of fit was performed under a balanced experiment's randomized complete block design (RCBD) (Table 5).

Table 5. Analysis of variance of the regression with the lack-of-fit test.

SOV	DF	SS	MS	F
Block	r - 1	SSBl	–	–
Regression	2	SSReg	SSReg/2	MSReg/ MSRegRes
RegRes	n - r - 2	SSRegRes	SQRegRes/(n - r - 2)	
Lack of Fit	n_{trat} - 3	SSLF	SSLF/ (n_{trat}- 3)	MSLF/MSRes
Residual	n - r - n_{trat} + 1	SSRes	SSRes/(n - r - n_{trat} + 1)	

Evaluated Measures

To compare, within each value of n (18, 24, 30, and 36), the different double factorial experiments, based on 1,000 simulations, were analyzed for the following four variables:

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

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

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

$$\mathrm{R} = \frac{1}{1.000}\sum_{s=1}^{1.000}\frac{\mathrm{SQReg}_s}{\mathrm{SQReg}_s+\mathrm{SQResReg}_s}.$$

The MAPEs (Mean Absolute Percentage Errors) show the absolute differences between the parameters and the estimates obtained by the respective fitted models of first-degree linear regression. For a perfect analysis, it would be expected that all of them are equal to zero. And for the measures R and ER, the higher their values, the better the fit of the first-degree linear regression model and the efficiency of RCBD, respectively.

In this study, the following factorial designs were compared:

2 x 3, 3 x 3 (FCCD) and RCCD, for n = 18 (Table 1);

DOI: 10.37856/bja.v98i3.4334

2 x 2, 2 x 3 e 3 x 4, for n = 24 (Table 2);

2^2 + 1, 2 x 3 e 3 x 5, for n = 30 (Table 3); and

2 x 2, 2 x 3, 3 x 3 (FCCD), 3 x 4, 3 x 6 and RCCD, for n = 36 (Table 4).

This means that 15 non-replicated double factorial experiments were generated based on the averages of 1,000 simulations and, according to the RCBD (randomized complete block design), distributed as follows: three experiments for n = 18, three for n = 24, three for n = 30, and six for n = 36. To evaluate them, combinations between the levels of two factors were established as follows:

c: number of combinations between the quantitative levels (treatments) of X_1 and X_2; and

d: RCCD, with 0 for no and 1 for yes.

For each of the four evaluated measures ($MAPE_{\beta 0}$, $MAPE_{\beta 1}$, $MAPE_{\beta 2}$, and R) and within the value of n = 36, a response surface analysis was performed to assess the effects of the number of combinations between the quantitative levels of X_1 and X_2 (c) and the RCCD (d), whose largest adopted model was defined as:

$$y_{ij}\ \beta_0 + \beta_1 c_i + \beta_2 d_j + \beta_3 c_i d_j + \varepsilon_{ij}, \text{ where:}$$

y_{ij}: the observed value of the measured variable at the combination of levels related to the number of combinations between the quantitative levels of X_1 and X_2, that is, at the values of c_i [4, 6, 9, 12, and 18 (n = 36)] and the values of d_j (0 and 1);

β_0: regression constant;

β_0, β_1, and β_3: regression coefficients; and

ε_{ij}~ N (0, σ_ε^2).

Subsequently, to fit the best model, at least one non-significant effect was removed one at a time, starting with the most complex to interpret, according to the Student's t-test at a 5% significance level.

For each of the four evaluated measures ($MAPE_{\beta 0}$, $MAPE_{\beta 1}$, $MAPE_{\beta 2}$, and R) within the values of n = 24 and n = 30, separately, a linear regression analysis was performed to assess only the effects of c, whose adopted model was defined as:

$$y_i = \beta_0 + \beta_1 c_i + \varepsilon_i \text{ where:}$$

y_{ij}: the observed value of the measured variable at the level related to the number of quantitative levels. (c_i) [4, 6, and 12 (n = 24), and 5, 6, and 15 (n = 30)];

β_0: regression constant;

β_1: regression coefficients; and

ε_i~ N (0, σ_ε^2).

For n = 18, only one descriptive statistic was performed for each measurement.

The statistical analyses conducted within each value of n (18, 24, 30, and 36) aimed to verify whether, for different double factorial experiments installed under the randomized complete block design (RCBD), it would be better to evaluate fewer combinations between the quantitative levels with more repetitions or more combinations between the quantitative levels with fewer repetitions, while considering the same number of experimental units (n) in a multiple linear regression analysis.

All simulations and statistical analyses related to the multiple linear regression model evaluations were performed using R version 4.0.2 (R CORE TEAM, 2020).

RESULTS AND DISCUSSION

For n = 18, the highest estimates of $MAPE_{\beta 0}$, $MAPE_{\beta 1}$, $MAPE_{\beta 2}$, were provided by the RCCD, indicating that it is likely to be less efficient in providing estimates closer to the parameters β_0, β_1, and β_2 when compared to the 2 × 3 and 3 × 3 designs (FCCD). And about the latter two, the 2 × 3 factorial design was slightly better (Figure 1).

Reinforcing the better performance of the 2 × 3 factorial design, we have the highest estimate of the mean of the evaluated measure R, which is related to the degree of explanation of the adjusted regression equation on the dependent variable Y (Figure 2). This confirms that the lower the number of combinations between the quantitative levels of X_1 and X_2 combined with the highest possible number of replications when planned in an experiment conducted under RCB, the better the quality of fit of the multiple linear regression analysis.

In this case, using the 2 × 3 factorial design is recommended instead of the RCCD and the 3 × 3 (FCCD). Similarly, Penteado and Batista (1971) obtained similar results when comparing the efficiencies of RCB and the 5 × 5 factorial design. In addition to them, Campos (1967) concluded that the 3 × 3 (FCCD) was more precise than RCCD, based on the estimates of the variances of the response surface estimators.

For n = 24 and n = 30 ($MAPE_{\beta 0}$), and n = 30 ($MAPE_{\beta 2}$), the means of the two evaluated measures decreased ($P < 0.05$) as the number of combinations between the quantitative levels of X_1 and X_2 (treatments) decreased. However, for $MAPE_{\beta 1}$ and $MAPE_{\beta 2}$ (n = 24), the estimated means did not change significantly ($P > 0.05$). On the other hand, the mean of the evaluated measure R increased ($P < 0.05$) as the number of treatments decreased, both for n = 24 and n = 30 (Table 6).

For all experiments conducted under RCBD with the same number of experimental units (n) equal to 36, the means of the evaluated measures $MAPE_{\beta 0}$ and $MAPE_{\beta 2}$ decreased ($P < 0.05$) as the number of combinations

Figure 1. Estimates of $MAPE_{\beta 0}$, $MAPE_{\beta 1}$, and $MAPE_{\beta 2}$ provided by RCCD, 2 x 3 e 3 x 3 (FCCD), for n = 18.

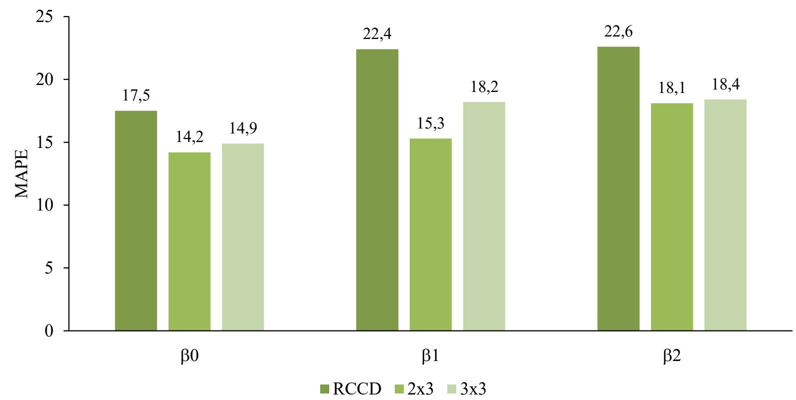
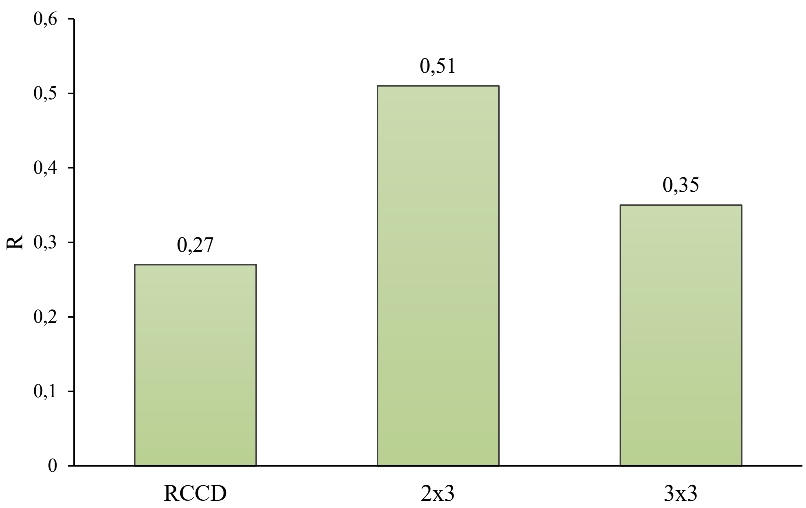


Figure 2.Estimates of R provided by RCCD, 2 x 3 e 3 x 3 (FCCD), for n = 18.



DOI: 10.37856/bja.v98i3.4334

Table 6.
Adjusted regression equations of $MAPE_{\beta 0}$, $MAPE_{\beta 1}$, $MAPE_{\beta 2}$, and R as a function of the number of combinations between the quantitative levels X_1 and X_2 for each value of n.

Measure	n	Regression equation	R^2
$MAPE_{\beta 0}$	24	10,6065 _ 0,2197*c	0,99
	30	-11,0025 + 4,0688*c	0,99
$MAPE_{\beta 1}$	24	14,05	–
	30	13,12	–
$MAPE_{\beta 2}$	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

**: significant by the Student´s t-test (P<0,05); c = number of combinations between quantitative of X_1 and X_2 [4 ≤ c ≤ 12 (n = 24) and 5 ≤ c ≤ 15 (n = 30)].*

Table 7.
Adjusted response surface of $MAPE_{\beta 0}$, $MAPE_{\beta 1}$, $MAPE_{\beta 2}$, and R as a function of the number of combinations between the quantitative levels and the RCCD n=36

	Response surface	R^2
$MAPE_{\beta 0}$	8,9625 + 0,1342*c + 2,1325*d	0,98
$MAPE_{\beta 1}$	11,9548 + 3,9573*d	0,67
$MAPE_{\beta 2}$	9,9135 + 0,3130*c + 3,0606*d	0,95
R	0,8093 - 0,0278*c - 0,1226*d	0,98

**: significant by the Student´s t-test (valor p<0,05); c = number of combinations between quantitative of X_1 and X_2 (4 ≤ c ≤ 18); d = rotatable central composite design (0 = no and 1 = yes).*

between the quantitative levels of X_1 and X_2 decreased and in the absence of the RCCD. However, for $MAPE_{\beta 1}$, the mean decreased ($P < 0.05$) only in the absence of the RCCD. On the other hand, the mean of the evaluated measure R increased ($P < 0.05$) as the number of treatments decreased and in the presence of the RCCD (Table 7).

Consequently, for the same value of n, the lower the number of combinations between the quantitative levels without using the RCCD, the smaller the estimates of the mean absolute deviations about the respective parameters $MAPE_{\beta 0}$, $MAPE_{\beta 1}$, and $MAPE_{\beta 2}$, as well as the absolute differences between the adjusted and true values of the dependent variable (Y). Therefore, the lower the number of quantitative levels combined with the highest possible number of replications, when planned in an experiment conducted under RCBD, the better the quality of fit of the multiple linear regression analysis.

Thus, it was concluded that among the evaluated double factorial experiments, the performance of multiple linear regression analysis was better when the number of combinations between the quantitative levels was smaller. The number of blocks was larger for the same number of experimental units [c=4 (2×2) and r = 6, c=5 (2^2 + 1) and r=6, c=4 (2×2) and r=8]. Consequently, if there is an expectation of fitting this model, it is recommended to experiment with only the levels corresponding to the lower and upper bounds of the interval of the independent variables X_1 and X_2. This means that the more repetitions (blocks) of the same quantitative level of X_1 and the same quantitative level of X_2 are performed, the better the analysis performance.

There is no need to evaluate more quantitative levels of X_1 and X_2, except for their respective lower (LI) and upper (LS) limits, to provide fewer distances between intermediate levels located within the evaluated interval of the independent variables. There is no need to evaluate any of the following quantitative levels of X_1 and X_2:

$LI_1 < X_1 < LS_1$; and

$LI_2 < X_2 < LS_2$.

On the other contrary, if there is no prior expectation of fitting a linear model, it is recommended to use nine quantitative treatments from a 3 × 3 factorial design (FCCD) and no more than that, with the levels of X1 and X2 defined as follows:

$x_{1_1} = LI_1$; $x_{1_2} = PC_1$; $x_{1_3} = LS_1$; and

$x_{2_1} = LI_2$; $x_{2_2} = PC_2$; $x_{2_3} = LS_2$, where:

PC_1 e PC_2: central points of X_1 (mean of LI_1 and LS_1) and X_2 (mean of LI_2 and LS_2).

CONCLUSIONS

For fitting a multiple linear regression model with two independent variables in an experiment conducted under RCBD, the quality increases with a decrease in the number of combinations between the quantitative levels (treatments) and an increase in the number of replications per combination. This implies that, for the same number of experimental units, it is recommended to use the minimum number of combinations between the quantitative levels of the two independent variables. If there is an expectation for the model with only linear effects, it is recommended to use only two quantitative levels per independent variable evaluated in a 2 × 2 factorial design. Otherwise, three levels are recommended, evaluated in a 3 × 3 factorial design.

REFERENCES

BAHRY, C. A.; VENSKE, E.; NARDINO, M.; FIN, S. S.; ZIMMER, P. D.; SOUZA, V. Q.; CARON, B. O. 2013. Características morfológicas e componentes de rendimento da soja submetida à adubação nitrogenada. **Revista Agrarian**, Pelotas, v. 6, n.21, p.281-288.

BONILLA, J. A. 1995. **Métodos quantitativos**, Belo Horizonte: Editora Líttera Maciel, 2th ed. 249 p.

BOX, G. E. P.; DRAPER, N. R. 1987. **Empirical model – building and response surfaces**, New York: John Wiley & Sons, 1st ed. 669 p.

CAMPOS, H. 1967. **Aspectos da aplicação das superfícies de resposta a ensaios fatoriais 33 de adubação**, Piracicaba: teste de doutorado – ESALQ/USP, 82 p.

DRAPER, N. R.; SMITH, H. 1998. **Applied regression analysis**, New York: John Wiley & Sons, 3rd ed. 693 p.

GOMES, F. P. 2009. **Curso de estatística experimental**, São Paulo: Livraria Universo Agrícola, 15th ed. 451 p.

HOFFMANN, R.; VIEIRA, S. 1983. **Análise de regressão: uma introdução à econometria**, São Paulo: Hucitec, 2th ed.379 p.

LIMA, P. C.; ABREU, A. R. 2000. **Delineamento e análise de experimentos**, Lavras: FAEPE, 1th ed. 45 p.

MATEUS, N. B.; BARBIN, D.; CONAGIN, A. 2001. **O delineamento composto central e sua viabilidade de uso em algumas áreas de pesquisa. Revista Acta Scientiarum,** v.23, n.06, p.1537-1546.

MONTGOMERY, D. C. 2009. **Design and analysis of experiments**, New York: John Wiley & Sons, 7th ed. 656 p.

DOI: 10.37856/bja.v98i3.4334

MONTGOMERY, D. C.; RUNGER, G. C. 2012. **Estatística aplicada e probabilidade para engenheiros**, Rio de Janeiro: LTC, 5th ed. 521 p.

MYERS, R. H.; MONTGOMERY, D. C.; A ANDERSON-COOK, C. M. 2009. **Response surface methodology: process and product optimization using designed** experiments, New Jersey: John Wiley & Sons, 3rd ed. 680 p.

PENTEADO, A. F.; BATISTA, L. B. 1971. Eficiência do ensaio composto central (Box) em comparação com os fatoriais completos de dois fatores. **Congresso Brasileiro de Ciência do Solo**, vol. 13.

POSSATO JUNIOR, O.; BERTAGNA, F. A. B.; PETERLINI, E.; BALERONI, A. G.; ROSSI, R. M.; ZENI NETO, H. 2019. Survey of statistical methods applied in articles published in Acta Scientiarum. **Acta Scientiarum. Agronomy,** vol. 41, p. 01-10. Available at: < https://doi.org/10.4025/actasciagron.v41i1.42641> Accessed on: May. 5, 2023.

SHIEH, G.; JAN, S. L. 2004. The effectiveness of randomized complete block design. **Statistica Neerlandica,** Neerdanda, vol. 58, p.111-124. Available at: < https://doi.org/10.1046/j.0039-0402.2003.00109.x> Accessed on: May. 4, 2023.

WERKEMA, M. C. C.; AGUIAR, S. 2006. **Análise de regressão: como entender o relacionamento entre as variáveis de um processo**, Belo Horizonte: Werkema Editora, 306 p.

ZEVIANE, W. M. 2011. **Manual de planejamento e análise de experimentos com R**, Curitiba: UFPR, 1th ed. 276 p.

Received in: Aug, 11, 2023.
Accepted in: Dec, 12, 2023.