Brazilian Journal of Agriculture **v.98, n.1, p. 52 – 67, 2023**
DOI: 10.37856/bja.v98i1.4329

PERFORMANCE OF SIMPLE LINEAR REGRESSION ANALYSIS UNDER A RANDOMIZED COMPLETE BLOCK DESIGN

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

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

ABSTRACT

In experiments conducted under a randomized complete block design, the fitting of the simple linear regression model can be performed under different combinations of the number of treatments and the number of replications. To determine the best combination, considering the same number of experimental units, it was concluded through a data simulation study that the quality of the fit increases when regression is performed in experiments with fewer treatments and more replications. Therefore, if linearity is expected, it is recommended to use two treatments for model fitting. Otherwise, three treatments are recommended. This applies to experiments with coefficients of variation between 10% and 30%.

Keywords: Treatments, Replications, Experimental precision.

INTRODUCTION

In agricultural sciences, many experiments are conducted with quantitative levels (treatments) under the randomized complete block design (RCBD) with four replications per treatment, and their responses are analyzed using the simple linear regression model. Before fitting this model, it is necessary to determine the dependent variable (Y) and the range of values for the independent variable (X) defined by the lower and upper limits.

The simple linear regression analysis has been used in over 31% of approximately 1200 articles published in the Acta Scientiarum Agronomy journal between 1998 and 2016, and experiments conducted in randomized complete blocks accounted for 43% of them. Moreover, in most cases, these experiments involved four to six replications per treatment.

This survey has also been verified by Possato et al. (2019). In these surveys, scientific articles related to research in Brazil on the crops of beans, corn, and soybeans were analyzed.

However, as reported by Montgomery (2009) and (2012), experimental designs may need to be conducted with different numbers of replications to achieve more appropriate precision and, perhaps, better cost-benefit ratios.

As mentioned earlier, the quantitative levels (treatments) should include both the lower and upper bounds and at least one of the intermediate levels, given that the objective is to fit a simple linear regression model that covers the entire range under study. This allows for great flexibility in choosing the number of treatments. Therefore, the same number of experimental units, can be divided into two main categories: a larger more significant of treatments with fewer replications per treatment and a smaller number of treatments with more replications per treatment.

Thus, the higher frequency of experimental designs with quantitative levels (treatments) analyzed using simple linear regression under the randomized complete block design with four replications per treatment is related to better experimental planning. This study aimed to evaluate the effects of the number of treatments and replications on the performance of the linear regression model with a single independent variable using simulated data from experiments conducted under the randomized complete block design.

MATERIAL AND METHODS

Regression parameters

The simple linear regression model that represented the functional relationship between the dependent variable (Y) and the independent variable (X) was given by:

$y_{ij} = 1.000 + 10x_i + \varepsilon_{ij}$, for $0 \leq X \leq 100$, where:

y_{ij}: observed value of the dependent variable Y at quantitative level x_i (i = 1, 2, ..., t) and block b_j (j = 1, 2, ..., r);

$\beta_0 = 1.000$: regression constant;

$\beta_1 = 10$:regression coefficient;

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

$\mu_i = 1.000 + 10x_{1_i}$: population mean of the dependent variable Y at quantitative level x_i; and

$\mu = 1.500$: overall population means of the dependent variable Y.

The regression parameters (β_0 e β_1) were defined based on the equation of simple linear regression fitted to the following variables evaluated in a soybean experiment: onset of maturity (Y) and nitrogen application rate (X_1). This experiment involved applying nitrogen during the reproductive phase of soybean, between growth stages R_1 (beginning of flowering) and R_6 (grains completely filling the pod cavity), and evaluating the onset of maturity at stage R_7 (BAHRY et al., 2013).

Data simulation

For obtaining the regression residuals (ε_{ij}), 1,000 simulations were conducted for each analyzed scenario, following a normal distribution with a mean of zero and a standard deviation σ_ε, where:

e_{ij}: regression residual associated with the observed value y_{ij} (i = 1, 2, ..., t e j = 1, 2, ..., r).

Initially, for the simulation realizations, the values of σ_ε were defined to provide residual coefficients of variation (CV_ε) of 10%, 20%, and 30%, according to the following expression:

$$\mathrm{CV}_\varepsilon = 100 \times \frac{\sigma_\varepsilon}{\mu} = 100 \times \frac{\sigma\varepsilon}{1.500}.$$

For the different simulation realizations, σ_ε values of 150, 300, and 450 were adopted, respectively. Consequently, the following normal distributions are associated with the regression errors:

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

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

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

Randomized complete block design

For comparison purposes, 25 experiments were generated using the Randomized Complete Block Design (RCBD) with 25 combinations of the number of treatments (quantitative levels of X ranging from zero to 100) (t) and the number of blocks (r), in order to provide the same number of experimental units (n = tr) equal to 12, 16, 20, 24, 28, and 32.

The block effects, considering the experiments installed under the RCBD with $CV_\varepsilon = 30\%$ as references, were defined in order to provide approximately the same block sum of squares (SSBl) that would yield significant effects ($f_{cal_{Bl}} \geq f_{tab_{Bl}}$) in all experiments with the same value of n. This was done considering $\alpha = 0{,}05$.

In this study, four experiments were generated under the RCBD for n = 12 (Table 1), three for n = 16 (Table 2), four for n = 20 (Table 3), six for n = 24 (Table 4), four for n = 28 (Table 5), and four for n = 32 (Table 6), where:

x_i: quantitative level of the independent variable X (i = 1, 2, ..., t); and

ω_j: effect of block b_j (j = 1, 2, ..., r).

Table 1. Quantitative levels and block effects of the four experiments installed under the RCBD with n = 12.

t = 2 e r = 6		t = 3 e r = 4		t = 4 e r = 3		t = 6 e r = 2	
x_i	ω_j	x_i	ω_j	x_i	ω_j	x_i	ω_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	–

DOI: 10.37856/bja.v98i1.4329

Table 2. Quantitative levels and block effects of the four experiments installed under the RCBD with n = 16.

t = 2 e r = 8		t = 4 e r = 4		t = 8 e r = 2	
x_i	ω_j	x_i	ω_j	x_i	ω_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	–

Table 3. Quantitative levels and block effects of the four experiments installed under the RCBD with n = 20.

t = 2 e r = 10		t = 4 e r = 5		t = 5 e r = 4		t = 10 e r = 2	
x_i	ω_j	x_i	ω_j	x_i	ω_j	x_i	ω_j
0	–352	0	–322	0	–288	0	–228
100	–277	33,33	–161	25	–144	11,11	228
–	–202	66,67	0	50	144	22,22	–
–	–127	100	161	75	288	33,33	–
–	–52	–	322	100	–	44,44	–
–	52	–	–	–	–	55,56	–
–	127	–	–	–	–	66,67	–
–	202	–	–	–	–	77,78	–
–	277	–	–	–	–	88,89	–
–	352	–	–	–	–	100	–

Table 4. Quantitative levels and block effects of the four experiments installed under the RCBD with n = 24.

t = 2 e r = 12		t = 3 e r = 8		t = 4 e r = 6		t = 6 e r = 4		t = 8 e r = 3		t = 12 e r = 2	
x_i	ω_j	x_i	ω_j	x_i	ω_j	x_i	ω_j	x_i	ω_j	x_i	ω_j
0	–318	0	–320	0	–302	0	–258	0	–250	0	–204
100	–263	50	–220	33,33	–177	20	–129	14,29	0	9,09	204
–	–208	100	–120	66,67	–52	40	129	28,57	250	18,18	–
–	–153	–	–38	100	52	60	258	42,86	–	27,27	–
–	–98	–	38	–	177	80	–	57,14	–	36,36	–
–	–43	–	120	–	302	100	–	71,43	–	45,45	–
–	43	–	220	–	–	–	–	85,71	–	54,55	–
–	98	–	320	–	–	–	–	100	–	63,64	–
–	153	–	–	–	–	–	–	–	–	72,73	–
–	208	–	–	–	–	–	–	–	–	81,82	–
–	263	–	–	–	–	–	–	–	–	90,91	–
–	318	–	–	–	–	–	–	–	–	100	–

Table 5. Quantitative levels and block effects of the four experiments installed under the RCBD with n = 28.

t = 2 e r = 14		t = 4 e r = 7		t = 7 e r = 4		t = 14 e r = 2	
x_i	ω_j	x_i	ω_j	x_i	ω_j	x_i	ω_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	–

Table 6. Quantitative levels and block effects of the four experiments installed under the RCBD with n = 32.

t = 2 e r = 16		t = 4 e r = 8		t = 8 e r = 4		t = 16 e r = 2	
x_i	ω_j	x_i	ω_j	x_i	ω_j	x_i	ω_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	–

Thus, for each of the 18 combinations between the values of $CV_{\varepsilon'}$ (10%, 20%, and 30%) and n (12, 16, 20, 24, 28, and 32), 1,000 simulations were performed according to their respective normal distributions ($\mu_\varepsilon = 0$ e σ_ε), in order to generate the n (tr) regression residuals.

Subsequently, the observed values of the dependent variable Y in each of the 25 balanced experiments installed under the RCBD were obtained as follows:

$y_{ij} = 1.000 + 10x_i + \omega_j + e_{ij}$, where:

y_{ij}: observed value of the dependent variable Y at quantitative level x_{1i} (i = 1, 2, ..., t) and block b_j (j = 1, 2, ..., r).

A total of 75 (3 × 25) different datasets were generated for the study of simple linear regression analysis, and for each of them, 1,000 simulations were performed.

For each of the 75,000 datasets, a first-degree linear regression model was fitted as follows:

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$, where:

$\hat{y}$: predicted value of the dependent variable Y at quantitative level x_i (i = 1, 2, ..., t) and block b_j (j = 1, 2, ..., r).

Subsequently, an analysis of variance for regression with a lack-of-fit test was performed under the RCBD of a balanced experiment (Table 7).

Table 7. Analysis of variance for regression with the lack-of-fit test.

SOV	DF	SS	MS	F
Block	r – 1	SSBl	–	–
Regression	1	SSReg	SSReg/1	MSReg/MSRegRes
RegRes	r(t – 1) – 1	SSRegRes	SSRegRes/[r(t – 1) – 1]	
Lack of Fit	t – 2	SSLF	SSLF/(t – 2)	MSLF/MSRes
Residual	(t – 1)(r – 1)	SSRes	SSRes/[(t – 1)(r – 1)]	

Evaluated measures

To compare, within each value of n (12, 16, 20, 24, 28, and 32), the effects of CV (in percentages, cv = 10%, 20%, and 30%) and the number of quantitative levels of X (t = 2, 3, 4, and 6 for n = 12, t = 2, 4, and 8 for n = 16, t = 2, 4, 5, and 10 for n = 20, t = 2, 3, 4, 6, 8, and 12 for n = 24, t = 2, 4, 7, and 14 for n = 28, and t = 2, 4, 8, and 16 for n = 32), the following five variables were analyzed based on the 1,000 simulations:

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

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

$$\mathrm{MAPE}_{\mu} = \frac{1}{1.000}\sum_{s=1}^{1.000} f_s;$$

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

$$ER = \frac{1}{1.000}\sum_{s=1}^{1.000}\frac{(r-1)\mathrm{QMBl}_s+r(t-1)\mathrm{QMResReg}_s}{(tr-1)\mathrm{QMResReg}_s}, \text{ where:}$$

$f_s = \frac{1}{16}\sum_{i=1}^{16}\left|\frac{\hat{y}_i - \mu_i}{\mu_i}\right| \times 100$, for x_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; and 100.

The MAPE (Mean Absolute Percentage Error) shows the absolute differences between the parameters and the estimates obtained from the respective fitted models of simple linear regression. For a perfect analysis, all values would be expected to be zero. And for the measures R and ER, the higher their values, the better the fit of the simple linear regression model and the efficiency of the randomized complete block design (RCBD), respectively.

As mentioned before, for n = 12 (Table 1), the effects of three CV_ε values and four quantitative levels (t = 2, 3, 4, and 6) were analyzed. For n = 16 (Table 2), the same $CV_{\varepsilon'}$ values were combined with three quantitative levels (t = 2, 4, and 8). For n = 20 (Table 3), with four quantitative levels (t = 2, 4, 5, and 10). For n = 24 (Table 4), with six quantitative levels (t = 2, 3, 4, 6, 8, and 12). For n = 28 (Table 5), with four quantitative levels (t = 2, 4, 7, and 14). And for n = 32 (Table 6), with four quantitative levels (t = 2, 4, 8, and 16).

This means that six factorial experiments were generated (3 × 4, 3 × 3, 3 × 4, 3 × 6, 3 × 4, and 3 × 4) were generated without repetitions for each combination of levels from the two factors (CV_ε e t), based on the means of the 1,000 simulations and following a completely randomized design (CRD).

For each of the five evaluated measures ($MAPE_{\beta_0}$, $MAPE_{\beta_1}$, $MAPE_\mu$, R e ER) in each of the 75 different data sets, a response surface analysis was conducted separately to assess the effects of the number of quantitative levels (t) and CV_ε (cv) levels for each number of experimental units (n). The adopted model for the analysis was defined as follows:

$y_{ij} = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \beta_3 b_j + \beta_4 b_j^2 + \beta_5 a_i b_j + \varepsilon_{ij}$, where:

y_{ij}: the observed value of the evaluated measure in the combination of levels related to the number of quantitative levels (a_i) [2, 3, 4, and 6 (n = 12), 2, 4, and 8 (n = 16), 2, 4, 5, and 10 (n = 20), 2, 3, 4, 6, 8, and 12 (n = 24), 2, 4, 7, and 14 (n = 28), and 2, 4, 8, and 16 (n = 32)] and CV_ε values (b_j) (10, 20 e 30);

β_0: regression constant;

β_1, β_2, β_3, β_4 e β_5: regression coefficients; and

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

Subsequently, to fit the best model, non-significant effects, if any, were removed one at a time, starting with the most complex one in terms of interpretation. If multiple effects had the same complexity, the effect with the highest p-value was removed as long as it was non-significant. However, non-significant effects that had a lower hierarchy compared to their respective significant effects were retained in the model.

The statistical analyses conducted within each value of n (12, 16, 20, 24, 28, and 32) aimed to investigate whether, for different experiments conducted under randomized complete block design (RCBD) with different precisions, it was better to evaluate fewer quantitative levels (treatments) with more replicates or more quantitative levels with fewer replicates, considering the same number of experimental units (n) in a simple linear regression analysis.

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

RESULTS AND DISCUSSION

Simple linear regression

For all experiments installed under the RCBD with the same number of experimental units (n), namely 12, 16, 20, 24, 28, and 32, the means of the evaluated measures $MAPE_{\beta_0}$, $MAPE_{\beta_1}$, and $MAPE_{\mu}$ decreased (p-value < 0,05) as the number of quantitative levels of X (treatments) and the residual coefficient of variation (CV_{ε}), in percentage, decreased (Table 8).

Consequently, for the same value of n, the smaller the number of quantitative levels (t = 2) and the CV_{ε} (cv = 10), the lower the mean absolute deviations of the estimates of β_0 and β_1 from their respective parameters, as well as the absolute differences between the adjusted values of the dependent variable (Y) and the respective parametric means conditioned on each level of the independent variable (X). Therefore, the smaller the number of quantitative levels combined with the highest possible number of repetitions planned in an experiment conducted under RCBD, the better the quality of fit of the simple linear regression analysis. Similarly, Mateus et al. (2001), when comparing coefficients of variation equal to 3%, 6%, 10%, 15%, and 24%, concluded that when the experimental coefficient of variation is more significant than 6%, it will be necessary to use more repetitions per treatment.

Table 8. Adjusted response surfaces of $MAPE_{\beta_0}$, $MAPE_{\beta_1}$, and $MAPE_{\mu}$ as a function of the number of quantitative levels and coefficient of variation for each value of n.

Measure	n	Response surface	R^2
$MAPE_{\beta_0}$	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
$MAPE_{\beta_1}$	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
$MAPEA_{\mu}$	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

*: significant by Student's t-test (p-value < 0,05); t = number of quantitative levels of X [$2 \leq t \leq 6$ (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), and $2 \leq t \leq 16$ (n = 32)]; cv = residual coefficient of variation, in percentage ($10 \leq cv \leq 30$).

Reinforcing the better-fit quality for experiments conducted under the RCBD with the same number of experimental units, when planned with smaller values of t and conducted under conditions with fewer uncontrollable effects, the average of the evaluated measure R increased (p-value < 0.05) due to the decrease in the number of quantitative levels and the residual coefficient of variation, in percentage. Consequently, there was a higher degree of explanation of X on Y, meaning that the adjusted model was closer to the true model. On the other hand, the average of the evaluated measure ER only increased (p-value < 0.05) due to the decrease in the coefficient of variation, except for n = 12 (Table 9).

Table 9. Adjusted response surfaces of R and ER as a function of the number of quantitative levels and coefficient of variation for each value of n.

Measure	n	Response surface	R^2
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

*: significant by Student's t-test (p-value < 0,05); t = number of quantitative levels of X [$2 \leq t \leq 6$ (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), and $2 \leq t \leq 16$ (n = 32)]; cv = coefficient of residual variation, in percentage ($10 \leq cv \leq 30$).

The higher the ER, the more efficient the CRD (completely randomized design) is compared to the RBD (randomized block design). According to the results, this efficiency increased when different blocks with t homogeneous experimental units were evaluated under lower occurrences of uncontrollable factors, that is, in experiments with lower residual coefficients of variation. It was concluded that the most significant relative reduction of residual variance (MSRegRes), the main advantage of using CRD instead of RBD (SHIEH; JAN, 2004), occurred in the presence of lower CVs. Furthermore, it was also observed that ER was not related to the number or size of the blocks.

Thus, it was concluded that among the evaluated experiments, the one with $t = 2$ and $r = 10$ ($n = 20$), $t = 2$ and $r = 12$ ($n = 24$), $t = 2$ and $r = 14$ ($n = 28$), and $t = 2$ and $r = 16$ ($n = 32$) exhibited the best performance in the simple linear regression analysis. Consequently, if there is an expectation of fitting this model, it is recommended to only experiment with the levels corresponding to the lower and upper bounds of the independent variable X. This means that the more repetitions (blocks) of the same quantitative level of X are conducted, the better.

In this regard, there is no need to evaluate additional quantitative levels of X, except for the lower bound (LB) and upper bound (UB), in order to provide fewer gaps between the intermediate levels within the assessed range of the independent variable. In fact, there is no need to evaluate any of the following quantitative levels of X:

$LB < X < UB$.

On the other hand, if there is no a priori expectation of fitting the linear model, it is recommended to use three quantitative treatments and no more than that, defined as follows:

$x1_1 = LB$;

$x1_2 = MC$; and

$x1_3 = UB$, where:

MC: mid-level (average of the lower and upper bounds).

In the linear relationship between Y and X_1, there are often different true models ($\mu_i = \beta_0 + \beta_1 x1_i$) and the adjusted models ($\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$). Therefore, as in the case of fitting a line, only two points (quantitative levels of X_1) are necessary. The results showed that estimating the means of these points with more repetitions was preferable. Conversely, the more points, i.e., the more quantitative levels of X with their respective estimated means, and consequently, with fewer repetitions, the worse the linear fit. This means that the more repetitions associated with estimating a mean, the less associated error will occur. To illustrate this conclusion, graphs of Y as a function of X were constructed based on the first of 1,000 simulations performed with $n = 32$ and $cv = 30\%$, for $t = 2$ and $r = 16$, and for $t = 16$ and $r = 2$ (Figure 1).

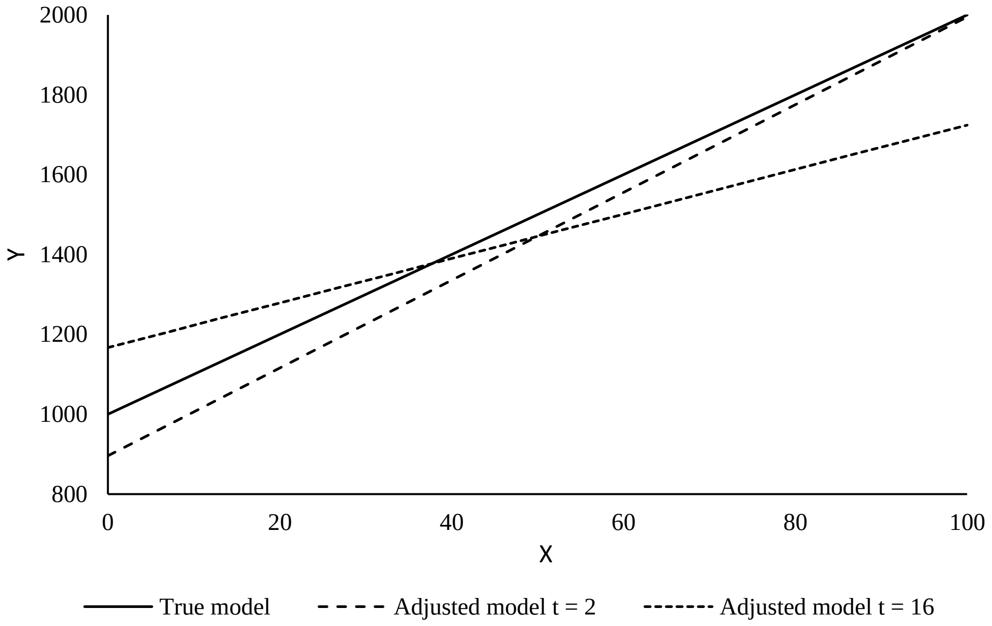

Figure 1. True linear models ($\mu_i = 1.000 + 10x_i$) and adjusted Y as a function of X for t = 2 and r = 16 ($\hat{y}_i = 896,5323 + 10,9832x_i$) and for t = 16 and r = 2 ($\hat{y} = 1.167,1042 + 5,5717x_{1_i}$).

Furthermore, as expected, the lower the $CV_{\varepsilon'}$, the better the fit of the regression equation and the more accurate the estimates of its parameters. The lower the CV_{ε}, the lower the residual variation (MSRegRes), and therefore, the higher the experimental precision. Consequently, treatment means will have estimates with smaller errors, even for those estimated with the smallest number of repetitions.

The CV_{ε}, when estimated, expresses the square root of the MSResReg as a percentage of the overall mean estimate of Y obtained in the experiment. Being dimensionless, it can provide an idea of the experiment's precision.

Under equal conditions, the experiment that provides a lower coefficient of variation will be more precise. However, the experiment with a higher number of repetitions per treatment is also expected to be more precise. Thus, to achieve the same precision in relation to the estimate of β_0 for experiments conducted under the CRD with n = 32, the following combinations can be used for example t = 2 (r = 16) and cv = 14.18, t = 8 (r = 4) and cv = 12.39, and t = 16 (r = 2) and cv = 10 (Table 8). For the estimate of β1, the following combinations are mentioned: t = 2 (r = 16) and cv = 17.77, t = 8 (r = 4) and cv = 14.44, and t = 16 (r = 2) and cv = 10 (Table 8).

CONCLUSIONS

For fitting a simple linear regression model in an experiment conducted under the RCBD, the quality improves with a decrease in the number of treatments and an increase in the number of replications per treatment. This implies that using the smallest possible number of treatments for the same number of experimental units is recommended. If there is an expectation for a linear model, using only two treatments (quantitative levels) is recommended. Otherwise, it is advisable to use a maximum of three treatments.

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, 1th ed. 669 p.

CAMPOS, H. 1967. **Aspectos da aplicação das superfícies de resposta a ensaios fatoriais 3^3 de adubação**, Piracicaba: teste de doutorado – ESALQ/USP, 82 p.

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

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.; ANDERSON-COOK, C. M. 2009. **Response surface methodology: process and product optimization using designed experiments**, New Jersey: John Wiley & Sons, 3th ed. 680 p.

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> Accesed 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> Accesed 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: May, 17, 2023.
Accepted in: June, 15, 2023.