

Article

Best Stand Correction Methods for Attenuating the Effects of Plant Loss in Experimental Plots of *Coffea arabica* Progenies

Cesar Elias Botelho ^{1,†}, Vinicius Teixeira Andrade ^{1,†}, Juliana Costa de Rezende Abrahão ^{1,*,†} and Flávia Maria Avelar Gonçalves ^{2,†}

¹ Southern Minas Regional Unit ,Agricultural Research Corporation of Minas Gerais (Epamig)/ Lavras, MG, Brazil cesarbotelho@epamig.br (CEB),vinicius.andrade@epamig.br (VTA) julianacosta@epamig.br (JCRA)

² Department of Biology, UFLA, Lavras, MG, Brazil avelar@ufla.br (FMAG)

* Correspondence: julianacosta@epamig.br.

† These authors contributed equally to this work.

Abstract: Plant loss in experimental plots occurs occasionally in field experiments with coffee crops. In breeding programmes, such loss can be extremely harmful, especially when the statistical analysis methods used are not consistent with the data generated in the experiments. In this study, we analysed a set of productivity data to determine whether the compensatory effect occurs in coffee crops, analyse the need for correcting the number of failures in experiments, and identify the best stand correction method to use. Productivity data from six harvests of 11 experiments with *Coffea arabica* plants were used. The experiments were implemented in a randomised block design, with four replications and six plants per plot. The following stand correction methods were evaluated: rule of three, Zuber [1], Vencovsky and Cruz [2] covariance of the average or ideal stands, and Cruz [3] and compared to data without correction adjustments. The most adequate correction methods were chosen based on the existence of genetic variance, selective accuracy, and progeny ordering. The compensatory effect was evident from the analysed data, with stand correction shown to be beneficial in progeny competition experiments. The covariance methods using average or ideal stands presented the best results, followed by the method proposed by Cruz [3]. The rule of three and Zuber [1] methods showed unsatisfactory results and are not recommended for stand correction in progeny competition experiments with coffee crops.

Keywords: *Coffea arabica*; stand correction; compensatory effect; unbalanced experiments

1. Introduction

The productivity performance of coffee crops in Brazil is linked mostly to the evolution of cultivation techniques and the use of improved cultivars [4]. Productivity gains have become significant over time and were achieved directly through the selection of cultivars with better production components and indirectly via the development of lineages resistant and/or tolerant to abiotic and biotic stresses.

In breeding programmes, progeny evaluation is the most costly and time-consuming stage. Experiments must be conducted with the greatest rigour and precision possible so that they have less errors and phenotypic differences represent the genotypic ones. The magnitude of experimental errors is directly related to the success of agricultural experimentation and plant breeding. Experimental accuracy is affected by several factors, such as the soil heterogeneity, presence of pests and diseases, plot size, design used, number of replications, and number of plants per plot [5].

There are many questions about the methods used in the evaluation of coffee progenies, and some of them have not yet been thoroughly researched. An example is plant loss in the plots, which reduces experimental precision and hinders identification of the best progenies [6]. Thus, measuring the effect of experimental plot failures and using methods that can attenuate such effect are necessary for improving the accuracy of estimates of genetic parameters.

Several stand correction approaches are found in the literature. A remarkably simple method is the rule of three (RT), which assumes proportionality between productivity and the number of plants in the experimental plot. However, such proportionality does not always occur, often resulting in a biased productivity value for the genotypes evaluated [2]. Zuber [1] proposed a second method (hereinafter referred to as 'Zb') to correct the RT error. However, this adjustment method also has limitations, as it does not consider the disposition of faults in the field, and its coefficient of compensation for lack of competition defaults to 0.3. Other methods which have proven to be efficient in many cases apply analysis of covariance, where the final stand of the plot is assumed as a covariate. The methods proposed by Cruz [3] and Vencovsky and Cruz [2] (hereinafter 'Cr' and 'VC', respectively) have also been used efficiently.

The compensatory effect of the crop under plot failure is another issue to be considered, since the plants adjacent to the failures generally have higher productivity owing to less competition for water, nutrients, and light. Although the compensatory effect is known to vary among different crops [6, 7] there are no published reports about its occurrence in *Coffea* plants and how it affects the accuracy of experiments with coffee crops.

Therefore, the objectives of this present study were to determine the magnitude of the compensatory effect in *Coffea* crops, to verify the need for correction of the number of failures in experiments, and to identify the best stand correction methods to use. The knowledge gained from this study will guide breeders on the appropriate data analysis procedures to use for plant selection in the event of stand reduction due to management and/or climatic events.

2. Materials and Methods

The analyses were performed using production data (in kilograms per plot) from six harvests obtained from 11 *Coffea arabica* progenies testing experiments conducted by the Agricultural Research Company of Minas Gerais (Empresa de Pesquisa Agropecuária de Minas Gerais - EPAMIG). The experiments, which were carried out in four cities in the state of Minas Gerais (Table 1), were performed using a randomised block design, with four replications and six plants per plot.

Table 1. List of experiments, progenies, locations, spacing, number of progenies (NP), percentage of failures (%), significance (p-value) for the number of failures in the plots, general mean ($\bar{Y}_{...}$), and linear regression coefficient (b).

t

Experiments	Progenies	Location	Spacing	NP	Failures (%)	p- value	$\bar{Y}_{...}$	b
IxC-CA	Icatu x Catimor	Campos Altos	3.5 x 0.5	30	6,80	0,16	19,30	1,11
IxC-S	Icatu x Catimor	São Sebastião do Paraíso	3.0 x 0.5	30	14,72	0,01		
MxC-TP	Mundo Novo x Catuaí	Três Pontas	2.5 x 0.7	42	22,88	0,28	12,03	1,62
MxC-CA	Mundo Novo x Catuaí	Campos Altos	3.5 x 0.5	25	4,33	0,31	25,66	-1,07
MxC-CP	Mundo Novo x Catuaí	Capelinha	3.5 x 0.5	25	12,83	0,89	11,11	0,88
ICT-CA	Icatu	Campos Altos	3.5 x 0.8	15	13,61	0,69	25,96	0,94
ICT-CP	Icatu	Capelinha	3.5 x 0.8	15	26,67	0,13	15,91	1,68
MN-TP	Mundo Novo	Três Pontas	3.0 x 1.0	35	10,33	0,40	22,22	1,99
MN-CA	Mundo Novo	Campos Altos	3.5 x 0.8	35	18,22	0,67	23,03	1,72
MN-CP	Mundo Novo	Capelinha	3.5 x 0.8	35	29,78	0,08	13,85	1,95
CAT-CP	Catuaí	Capelinha	3.5 x 0.5	20	46,25	0,01		
Mean							19,15	1,20

F4 progenies from Icatu x Catimor and Mundo Novo x Catuaí crosses and lineages from the Mundo Novo, Catuaí, and Icatu groups were analysed (Table 1). Harvesting of the individual plots was conducted annually, with the coffee (coffee fruits of mixed maturity) volume obtained in each plot converted into 60 kg bags of processed coffee per hectare (bags.ha⁻¹).

The compensatory effect was determined by estimating the compensation coefficient (a) using the estimator $a=c/(\bar{Y}^-)$, where c is the linear regression coefficient proposed by Cruz [1] and $\bar{Y}^-$ is the

mean production yield per plant obtained in the experiment. An a -value equal to or greater than 1 indicates a positive compensatory effect on the crop.

Following the recommendation by Steel et al. [9], correction was not applied whenever significant differences in the final number of plants per plot (stand) between treatments were detected with the F test. In the absence of grain production data adjustments (AA), plant loss in the plot was disregarded (in this case, $Z_{ij} = Y_{ij}$). Otherwise, the productivity data were adjusted as a function of the variable stand, using the following six procedures: (i) the RT method, using the expression $Z_{ij} = Y_{ij}(H/X_{ij})$, where H represents the ideal stand (in this case, six plants); (ii) the Zb method, using $Z_{ij} = Y_{ij}[(H - a)(H - X_{ij})]/X_{ij}$, where a represents the compensation coefficient for the absence of competition (in this case, $a = 0.3$); (iii) the VC method, using $Z_{ij} = Y_{ij}[(H - a)(H - X_{ij})]/X_{ij}$, where a represents the compensation coefficient estimated from experimental data; (iv) analysis of covariance with the ideal stand as the covariate (hereinafter 'ACI' method), using $Z_{ij} = Y_{ij} - b(X_{ij} - H)$, where b represents the linear regression coefficient as a function of Y_{ij} , estimated according to the procedure described by Steel et al [1]; (v) analysis of covariance with the average stand as the covariate (hereinafter 'ACA' method), using $Z_{ij} = Y_{ij} - b(X_{ij} - A)$, where b represents the Y_{ij} residual regression coefficient, estimated according to the procedure described by Steel et al [9] and A represents the average stand of the experiment; and (vi) the Cr method, using $Z_{ij} = Y_{ij}(H/X_{ij}) - c(H - X_{ij})$, where c represents the residual regression coefficient of the variable Y_{ij} (corrected using the RT method) as a function of the number of failures in the plot. In all the adjustment expressions above, Z_{ij} represents the corrected yield and Y_{ij} the observed production in real plots/stands (X_{ij}).

Analyses of the six harvests were carried out using Z_{ij} data in a split-plot design over time [9] by applying the following statistical model: $y_{ijq} = \mu + b_j + p_i + c_q + [pb]_{ij} + [bc]_{jq} + [pc]_{iq} + e_{ijq}$, where y_{ijq} represents the observation of the ijq -th plot in block j of harvest q which received progeny i ; μ represents the constant associated with all observations; b_j is the fixed effect of the j -th block; p_i is the random effect of the i -th progeny, where $p_i \sim \text{NMV}(0, \sigma_p^2)$; c_q is the fixed effect of the q -th harvest; $[pb]_{ij}$ is the random effect of the ij -th progeny-block interaction, where $[pb]_{ij} \sim \text{NMV}(0, \sigma_{pb}^2)$; $[bc]_{jq}$ is the fixed effect of the jq -th block-harvest interaction; $[pc]_{iq}$ is the random effect of the iq -th progeny-crop interaction, where $[pc]_{iq} \sim \text{NMV}(0, \sigma_{pc}^2)$; and e_{ijq} is the random effect of the experimental error associated with the observation of the ijq -th plot, where $e_{ijq} \sim \text{NMV}(0, \sigma_e^2)$. The empirical best linear unbiased prediction (E-BLUP) value was estimated by considering each adjustment method and the uncorrected data [10]. Estimates of the experimental accuracy were obtained algebraically using the square root of the heritability at the mean progeny level.

The magnitudes of the parameter estimates were used to verify which method would be the most efficient, including the mean ($\bar{Y}_{\dots}$), standard deviation of the mean (s), p -value, and selective accuracy ($r^2(g \hat{g})$). The most effective methods for stand correction are those that result in less interference in the mean, smaller s -values, smaller p -values, and greater $r^2(g \hat{g})$ estimates.

The E-BLUP estimates obtained with each correction method were used to estimate the index of coincidence (IC) for a selection intensity of 20%; that is, the proportion of superior progenies and/or lineages with the same behaviour for each method compared with the unadjusted data. The IC was estimated using the method proposed by Hamblin and Zimmermann [11], which considers the effect of chance: $IC = (A - C)/(M - C) \times 100$, where C represents the number of progenies and/or superior lineages selected as a result of chance (i.e. the number of selected progenies and/or superior lines in proportion to the selection intensity is assumed to coincide by chance); A is the number of superior progenies selected, common to the different methods; and M is the number of superior progenies and/or lineages selected with the AA method.

3. Results

Analysis of variance testing of the number of failures in each plot revealed whether they had occurred by chance (Table 2). The percentage of failures per experiment ranged from 4.33% to 46.25%, with the greatest loss being observed in the lineage of the Catuaí group. In two cases (Icatu $\times$ Catimor progenies in São Sebastião do Paraíso and Catuaí lineages in Capelinha), significant differences

between progenies with regard to plot stands were detected, indicating that the progeny effect influenced plant loss in the plots.

Table 2. List of experiments, sites, and compensation coefficient (a) values for experiments and harvests from 1 to 6 (H 01 to H 06).

Experiments	H 01	H 02	H 03	H 04	H 05	H 06	Mean
IxC-CA	0,89	1,38	1,61	1,21	-0,08	1,72	1,12
MxC-TP	2,77	1,77	2,86	0,56	1,72	0,07	1,63
MxC-CA	0,94	1,017	2,73	2,14	0,80	0,74	1,39
MxC-CP	-0,24	1,17	-0.16	2,66	0,99	0,99	1,00
ICT-CA	0,39	1,11	0,42	1,16	0,14	1,75	0,83
ICT-CP	0,43	1,48	0,54	2,40	1,25	2,04	1,35
MN-TP	-0,53	0,001	1,78	0,40	6,60	0,53	1,29
MN-CA	0,19	0,88	1,10	1,07	1,10	0,002	0,73
MN-CP	-0,05	1,04	-0,09	0,94	1,75	0,67	0,71
Avarege	0,53	1,09	1,37	1,39	1,59	0,95	1,15

Estimates of the mean linear regression coefficient (b), which reflects the coffee production response in the plot as a function of the change in the number of plants (stand), were generally expressive of the six harvests (Table 1). The experiment with Mundo Novo × Catuaí de Campos Altos progenies presented a negative b estimate, indicating that each failure resulted in a decrease of 1.07 kg in the plot. For the other cases, the b estimates ranged from 0.88 to 1.99, indicating an increase of 3.64–14.08% relative to the mean production in the plot.

Considering the existence of b, the compensation coefficient for lack of competition (a) was estimated for each experiment and year of harvest (Table 2). There was variation between groups of progenies, between locations, and (mainly) between harvesting years. The greatest variation occurred in the Mundo Novo lineage experiment in Três Pontas, where estimates ranged from −0.001 (H 02) to 6.60 (H 05). In general, compensation within the same experiment increased over time until a level of equilibrium was reached. Comparisons of the coefficients between experiments showed that the failures in the Mundo Novo × Catuaí experiment (conducted in Três Pontas) were the ones that changed the productive response pattern of the plants the most ($Y_{...}^- = 1.63$), since this experiment showed greater compensation for the lack of competition.

Having determined that failures did indeed change the productive response pattern of the coffee plants, it was necessary to apply methods that result in adequate selection. With this objective in mind, six methods for correcting the production data of the *Coffea arabica* progenies were tested. Comparisons of the adjusted data with the original data (AA) showed that the six correction methods used generally overestimated the means to different degrees (Table 3). The RT method provided the highest mean estimate, overestimating it by 18.72% in relation to the original data (AA). Likewise, the Zb method overestimated the mean of all trials by 15.56%. By contrast, the best mean estimates were obtained with the VC and ACA methods, which overestimated the unadjusted mean by only 4.34% and 4.72%, respectively (Table 3).

Table 3. General means of the experiments according to the various data correction methods used and their percentage increment in relation to the unadjusted mean.

Experiments *	AA	RT	Zb	VC	ACI	ACA	Cr
IxC-CA	19,30	20,98	20,47	20,98	19,75	19,30	20,98
MxC-TP	12,30	18,65	18,38	17,81	18,18	17,75	17,98
MxC-CA	25,67	27,11	26,68	24,80	25,36	25,67	25,49
MxC-CP	11,12	13,07	12,48	11,03	11,79	11,12	11,57
ICT-CA	25,95	31,27	29,68	29,09	26,73	25,96	25,89
ICT-CP	15,91	23,99	21,57	10,20	18,59	18,59	15,91
MN-TP	22,22	23,35	23,01	22,53	22,69	22,22	22,87
MN-CA	26,03	30,63	29,25	28,06	27,31	26,03	26,47
MN-CP	13,85	19,29	17,66	15,34	16,75	13,85	16,15
Average	19,15	23,15	22,13	19,98	20,79	20,05	20,36
Increment (%)	-	18,72	15,57	4,34	8,59	4,72	6,32
Standart deviation of the mean	4,16	6,81	5,57	5,64	3,95	3,95	5,57

AA: absence of adjustment; RT: rule of three method; Zb: Zuber [1] method; VC: Vencovsky and Cruz [2] method; ACI: analysis of covariance based on the ideal stand; ACA: analysis of covariance based on the average stand; Cr: Cruz [3] method. *IxC-CA: Icatu × Catimor in Campos Altos; IxC-S: Icatu × Catimor in São Sebastião do Paraíso; MxC-TP: Mundo Novo × Catuaí in Três Pontas; MxC-CA: Mundo Novo × Catuaí in Campos Altos; MxC-CP: Mundo Novo × Catuaí in Capelinha; ICT-CA: Icatu in Campos Altos; ICT-CP: Icatu in Capelinha; MN-TP: Mundo Novo in Três Pontas; MN-CA: Mundo Novo in Campos Altos; MN-CP: Mundo Novo in Capelinha; CAT-CP: Catuaí in Capelinha.

The standard deviation, which indicates the degree of data dispersion in relation to the mean, had the the highest value estimated in the RT method (6.81). This inferred that in addition to overestimating the data and not considering compensation, this method is inappropriate for increasing dispersion around the mean (Table 4).

Table 4. Genetic parameter estimates obtained by different methods used to correct production data from F4 progenies obtained by crossing cultivars of Icatu × Catimor, Mundo Novo × Catuaí, and Icatu in the cities of Campos Altos (CA), Capelinha (C), and Três Pontas (TP).

Icatu x Catimor - CA					Mundo Novo x Catuai -CA				Mundo Novo x Catuai- C			
	p-valor	$\hat{r}_{gg}$	r_G	IC	p-valor	$\hat{r}_{gg}$	r_G	IC	p-valor	$\hat{r}_{gg}$	r_G	IC
AA	0,09	0,61	1,00	1,00	0,12	0,60	1,00	1,00	0,29	0,42	1,00	1,00
RT	0,23	0,46	0,66	0,14	0,26	0,46	0,86	0,73	-	0,00	0,00	0,00
Zb	0,18	0,52	0,81	0,71	0,21	0,51	0,92	0,73	-	0,00	0,00	0,00
VC	0,14	0,55	0,95	0,71	0,10	0,62	0,92	1,00	0,20	0,52	0,97	0,73
ACI	0,14	0,53	0,98	1,00	0,09	0,64	0,99	1,00	0,45	0,21	0,98	0,73
ACA	0,12	0,56	0,97	1,00	0,09	0,64	0,99	1,00	0,45	0,21	0,98	0,73
Cr	0,08	0,63	1,00	1,00	0,09	0,64	0,99	1,00	0,40	0,28	0,82	0,47
Mundo Novo x Catuai - TP					Icatu - CA				Icatu - C			
	p-valor	$\hat{r}_{gg}$	r_G	IC	p-valor	$\hat{r}_{gg}$	r_G	IC	p-valor	$\hat{r}_{gg}$	r_G	IC
AA	0,07	0,56	1,00	1,00	0,33	0,43	1,00	1,00	0,04	0,82	1,00	1,00
RT	0,06	0,58	0,93	0,00	-	0,00	0,00	0,00	0,19	0,59	0,59	0,29
Zb	0,06	0,57	0,96	0,00	0,27	0,27	0,65	0,53	0,14	0,65	0,71	0,29
VC	0,04	0,56	0,99	0,66	-	0,00	0,00	0,00	0,24	0,54	0,38	0,53
ACI	0,06	0,58	0,98	0,66	0,33	0,43	0,98	1,00	0,04	0,82	0,96	0,76
ACA	0,06	0,58	0,98	0,66	0,33	0,43	0,98	1,00	0,04	0,82	0,96	0,76
Cr	0,05	0,58	0,99	0,66	0,41	0,30	0,90	0,76	0,14	0,65	0,71	0,53
Mundo Novo- CA					Mundo Novo -TP				Mundo Novo – CA			
	p-valor	$\hat{r}_{gg}$	r_G	IC	p-valor	$\hat{r}_{gg}$	r_G	IC	p-valor	$\hat{r}_{gg}$	r_G	IC
AA	0,08	0,61	1,00	1,00	0,22	0,46	1,00	1,00	0,02	0,72	1,00	1,00
RT	-	0,00	0,00	0,00	0,19	0,48	0,81	0,38	-	0,00	0,00	0,00
Zb	-	0,00	0,00	0,00	0,18	0,49	0,90	0,69	0,34	0,33	0,79	0,69
VC	0,42	0,23	0,84	0,08	0,47	0,15	0,95	0,69	0,09	0,59	0,96	1,00
ACI	0,20	0,48	0,96	1,00	0,17	0,50	0,97	0,69	0,06	0,64	0,92	1,00
ACA	0,20	0,48	0,96	1,00	0,17	0,50	0,97	0,69	0,06	0,64	0,92	1,00
Cr	0,11	0,57	0,93	0,69	0,21	0,47	0,91	0,69	0,31	0,36	0,84	0,69

p-value: probability of significance of the genetic factor; r_{gg} : selective accuracy; r_G : genetic correlation; IC: index of coincidence; AA: absence of adjustment; RT: rule of three method; Zb: Zuber [1] method; VC: Vencovsky and Cruz 2] method; ACI: analysis of covariance based on the ideal stand; ACA: analysis of covariance based on the average stand; Cr: Cruz [3] method.

For each set of adjusted data, the p-value, selective accuracy, and genetic correlation between progeny ordering and IC value of the five best progenies selected were estimated to evaluate the influence of plant loss on the selection carried out by coffee breeders in their routine work. The p-value was interpreted as a gradation of probabilities of accepting or rejecting H0, rather than being a fixed value above which no significant difference is considered. Therefore, regardless of the fixed level of significance, we analysed the possibility of a change in the variation probability due to the analysed factor.

Analysis of the F4 progenies generated by the Icatu × Catimor cross in Campos Altos showed that the lowest p-values were obtained with the Cr (0.08) and AA (0.09) methods, albeit they were not significant at the 5% level according to the standard procedure. These two methods were more accurate (0.63 and 0.61) and generated a 100% IC in the ACI and ACA methods (Table 4). Based on the IC values, the RT method resulted in the most discrepant transformation in the selections made (0.14). For this population, the best parameters were obtained using the Cr method, which resulted in greater selective accuracy and genetic correlation and an IC equal to the AA data, yielding a probability of differences due to genotype of 6.32%.

The progenies of Mundo Novo × Catuaí in Campos Altos obtained the best fit with the ACI, ACA, and Cr methods, which yielded the highest selective accuracy. About the correlation and

coincidence in selection, the ACI, ACA, and Cr methods again provided the same ordering and selection as the AA method. In Capelinha, the genetic variance was not significant, indicating the difficulties in selecting superior genotypes in that location. The VC method performed well, since it promoted the best genetic discrimination of progenies, as verified by the p-value. Thus, the correlation and IC values of the VC method were equal to those of the correction methods with the best fit (i.e. ACI and ACA), leading to the conclusion it may be the best data adjustment approach in this situation. Likewise, in Três Pontas, the VC method achieved greater discriminatory power and selected the same progenies as the methods with best fit (ACI and ACA) did (Table 4).

For the Icatu progenies, whether in Campos Altos or in Capelinha, the best parameter estimates were obtained with the ACI, ACA, and AA methods. There was no genetic variation in the progenies in Campos Altos, whereas genetic effects were highly significant in those in Capelinha and the selective accuracy was high. The genetic correlation of the ACI- and ACA-corrected data with the AA data was high, generating a selection IC of 100% in Campos Altos and 76% in Capelinha.

About the Mundo Novo progenies, the ACI and ACA methods stood out in fitting the data to the model in the three experimental locations. In Campos Altos, the AA method presented the highest selective accuracy value. An identical pattern was observed for the progenies in Capelinha. By contrast, the AA, ACI, and ACA methods were not able to identify genetic differences in the progenies in Três Pontas and selected only 70% of coinciding progenies.

4. Discussion

For perennial crops such as coffee, the breeder must pay attention to the use of appropriate experimental techniques that minimise experimental errors. Losses caused by the inadequacy of techniques will only be verified after several years and may reduce selection efficiency and delay the obtention of progenies and/or new cultivars. Regardless of whether they are controllable or not, all factors that affect the error must be observed by breeders in order to increase the efficiency of the selection process.

The loss of plants in the plot causes major problems in data analysis and experimental result interpretation, since it results in uneven stands and compromises experimental precision, with the consequences being greater for perennial crops as they perpetuate errors throughout the duration of the experiment [12]. Currently, information on how to deal with plant losses in plots of coffee experiments in the field remains scarce. Therefore, in this study, an extensive EPAMIG database was analysed to understand the influence of the compensatory effect of *Coffea arabica* progeny production on productivity data and identify alternatives for overcoming this problem.

In this study, the linear regression coefficient of productivity with the number of plants in the plot and the compensatory effect of the production of adjacent plants against an absent plant were estimated. This balance between production loss and compensation is genetically determined and depends on the planting spacing and arrangement failures in the plot [2]. Of the 54 harvests studied, 32 had a compensation coefficient value of greater than 1, indicating a positive mean capacity for production compensation as a function of each failure present in the plot of 1.15 (Table 3). Considering that this set of data is quite representative of actual scenarios, coffee trees possibly have a significant capacity to compensate for the absence of plants in the plot or a decrease in the stand in the case of wider spacing. These results appear to make biological sense, since compensation follows the natural productive increase in coffee trees, which reaches a peak between the sixth and eighth harvest, stabilising their biennial production until ageing of the plants [13]

The results obtained corroborate those of other researchers, who reported that plant production compensation in the same crop varies between experiments, demonstrating the influence of genotypes and soil and climate conditions on this parameter. These effects were verified in maize [15] and beans [16, 17]. In the present study, the observed differences in compensation for lack of competition (Table 3) can be attributed to the fact that coffee progenies vary in their ability to respond to spacing via root growth differentials [17] and respond to light through different plagiotropic branches, angles, and plant heights [18].

After verifying the fit of the model and analysing whether the variation is heritable and the phenotype faithfully represents the genotype, we observed the impact of data transformations in terms of plant loss in the plot on progeny selection. First, the analysis was carried out in a split-plot design over time, which is a common procedure for production data in coffee breeding as it identifies precocious progenies in relation to production and allows estimation of genetic variances free of progeny-crop and progeny-block interaction variances [19].

For progeny and cultivar competition tests, it is necessary to work on the data in order to verify their need for correction due to plant loss in the plots. Genetic correlation and IC analyses were then performed to determine the magnitude of progeny ordering change and the proportion of similarity between the selections carried out by the different methods in association with a probability chosen by the breeder. In this study, the most adequate methods for correcting the productive data of the progeny tests were chosen on the basis of the p-value and selective accuracy and verified for each population (Table 4). The results revealed that generalisations are not appropriate, making it impossible to establish fixed rules, since the interaction of the evaluated plant with the cultivation environment results in a relatively plastic response.

The estimated compensation coefficients obtained from the various experiments varied greatly when the fixed factor of 0.3 used in the Zb method was applied, revealing this method to be ineffective. Only 13% of the studied estimates were similar to the 0.3 value recommended by Zuber [1], indicating the need to reconsider this correction method for coffee crops.

Likewise, stand correction using the RT method was not effective, as it did not provide greater estimates of selective accuracy in any of the experimental situations studied and instead presented the highest production means, which overestimated the mean production value. According to Vencovsky and Cruz [2], this method has been efficient in situations where the degree of failure compensation in the plot was low but has proven to be inefficient in cases where the compensation was high (close to 1). Although the RT method is not useful for data correction, it is necessary for calculating the compensation coefficient due to irregularities in the experimental stand.

In general, the covariance correction methods for average and ideal stands were efficient, as they promoted greater decrease in the residual variance through a reduced p-value and increased the experimental precision, corroborating the findings by other researchers who analysed maize [14] and grain sorghum [17].

In the search to establish a relationship between selection decisions based on data corrections and those based on the original data (unadjusted), it was found that six of the nine selection decisions by the breeders were completely coincident with what was conducted in the populations of Icatu × Catimor [21,22], Mundo Novo × Catuaí [23,24], and lineages from the Mundo Novo [25], Catuaí [26], and Icatu groups [27]. Therefore, assuming that the adjustments are biologically correct, the breeders were correct 66.67% of the time. As breeding requires intensive use of resources, a 33.33% probability of error in selection is considerably high and demonstrates, once again, the importance of considering the effect of plant loss in the plot.

If we ignore the importance of properly applying a correction method in experiments with plot failures, the possible effects of competition between plots may result in the selection of genotypes that have limited value for the trait and target environments. Furthermore, many of the actual genetic effects are small in magnitude for a polygenic trait, potentially contributing to statistical bias and error [28].

Despite not being the focus of this study, data correction needs to be understood in the context of genotype-environment interaction. In this study, the data were analysed according to experimentation locations within populations. However, after identifying the most suitable methods for correcting failures, these data can be used in joint analyses of the environments under study, increasing the chance of correct decisions being made.

Another aspect identified through the interpretation of the results of this study is the criterion for choosing the correction methods to be applied. This decision should consider several parameters together and first analyse the existence of genetic variance, selective accuracy, correlation between phenotype and genotype, and ordering of the progenies, lineages, or cultivars under evaluation. The comparison of progenies under equal conditions is essential for ensuring the genetic progress of

cultivars and the efficiency of breeding programmes. Therefore, it is important to carry out experiments that allow plants to express their genetic potential. Although many unforeseen events may occur, the experiments need to be carefully conducted, especially in the crop formation phase.

Author Contributions: CEB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. VTA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology. JCRA: Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. FMAG Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding: This work was supported by Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG). The authors would like to thank for the financial support of Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Consórcio de Pesquisa Café, and Instituto Nacional de Ciência e Tecnologia do Café (INCT-Café).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zuber, M.S. Relative Efficiency of Incomplete Block Designs Using Corn Uniformity Trial Data. *Agronomy Journal* **1942**, 34(1), 30–47. doi:10.2134/agronj1942.00021962003400010004x.
2. Vencovsky, R.; Cruz, C.D. Comparação de métodos de correção do rendimento de parcelas com estandes variados. I. Dados simulados. *Pesquisa Agropecuária Brasileira* **1991**, 26(5), 647–657.
3. Cruz, V.F. Estudo sobre a correção de produção de parcelas em ensaios com milho. *Doctoral dissertation*, Escola Superior de Agricultura Luiz de Queiroz, São Paulo State University Digital Repository, **1971**.
4. van der Vossen, H.; Bertrand, B.; Charrier, A. Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. *Euphytica* **2015**, 204(2), 243–256. doi:10.1007/s10681-015-1398-z.
5. Kryszczun, D.K.; Lúcio, A.D.; Sari, B.G.; Diel, M.I.; Olivoto, T.; Santana, C.S.; Ubessi, C.; Schabarum, D.; Esequiel. Sample size, plot size and number of replications for trials with *Solanum melongena* L. *Scientia Horticulturae* **2018**, 233, 220–224. doi:10.1016/j.scienta.2018.01.044.
6. Mackay, I.; Piepho, H.P.; Garcia, A.A.F. Statistical methods for plant breeding. In *Handbook of Statistical Genomics: Two Volume Set*; John Wiley & Sons, **2019**, pp. 501–520.
7. McGinty, J.; Morgan, G.; Mott, D. Cotton Response to Simulated Hail Damage and Stand Loss in Central Texas. *Journal of Cotton Science* **2019**, 23(1), 1–6. doi:10.56454/OXUC6037.
8. Joukhadar, I.; Walker, S. Effect of Stand Reduction at Different Growth Stages on Yield of Paprika-Type Chile Pepper. *Horticulturae* **2020**, 6(1), 16. doi:10.3390/horticulturae6010016.
9. Steel, R.G.D.; Torrie, J.H.; Dickey, D.A. *Principles and procedures of statistics: a biometrical approach*, 3rd ed.; McGraw-Hill, **1997**.
10. SAS Institute Inc. SAS 9.4/STAT 15.1 *User's Guide*. Cary, NC: SAS Institute Inc., **2022**.
11. Hamblin, J.E.; Zimmermann, M.J.O. Breeding common bean for yield in mixtures. *Plant Breeding Reviews* **1986**, 4, 245–272.
12. Andrade, H.B.; Ramalho, M.A.P.; Bueno Filho, J.S.S.; Resende, M.D.V.; Xavier, A. Alternativas para atenuar a diferença de estande nos experimentos de avaliação de clones de *Eucalyptus urophylla*. *Revista Árvore* **2006**, 30(1): 11–18. doi: 10.1590/S0100-67622006000100002
13. Carvalho, A., Antunes Filho, H., Mendes, J. E. T., Lazzarini, W., Junqueira Reis, A., Sobrinho, J., & Rocha, T. D. Melhoramento do cafeeiro: XIII-Café bourbon amarelo. *Bragantia*, **1957** 16, 411–454.
14. Veronesi, J.A., Cruz, C.D., Corrêa, L.A., & Scapim, C.A. Comparação de métodos de ajuste do rendimento de parcelas com estandes variados. *Pesquisa Agropecuária Brasileira*, **1995** 30(2), 169–174.
15. Adams, M.W. Basis of Yield Component Compensation in Crop Plants With Special Reference to the Field Bean, *Phaseolus vulgaris*. *Crop Science*, **1967** 7(5), 505–510. doi: 10.2135/cropsci1967.0011183X000700050030x.
16. Ribeiro, N.D., Cargnelutti Filho A., Hoffmann Júnior L., & Possebon S.B. Precisão experimental na avaliação de cultivares de feijão de diferentes hábitos de crescimento. *Ciencia Rural*, **2004** 34(5), 1371–1377. doi: 10.1590/S0103-84782004000500008.
17. Silva, L.O.E., Schmidt R., Valani G.P., Ferreira A., Ribeiro-Barros A.I., & Partelli F.L. Root Trait Variability in *Coffea canephora* Genotypes and Its Relation to Plant Height and Crop Yield. *Agronomy Journal*, **2020** 10(9), 1394. doi: 10.3390/agronomy10091394.
18. Dubberstein, D., Partelli F.L., Guillen J.H.S., Rodrigues W.P., Ramalho J.C., & Ribeiro A.I. Biometric traits as a tool for the identification and breeding of *Coffea canephora* genotypes. *Genetic Molecular Researcher*, **2020** 19(2). doi: 10.4238/gmr18541.

19. Andrade, V.T., Gonçalves F.M.A., Nunes J.A.R., & Botelho C.E. Statistical modeling implications for coffee progenies selection. *Euphytica*, **2016**, 207, 177-189. doi: 10.1007/s10681-015-1561-6.
20. Silva, K.J. da, Menezes C.B. de, Tardin F.D., Souza V.F., & Santos C.V. Comparação de métodos de correção de estande para estimar a produtividade de sorgo granífero. *Pesquisa Agropecuaria Tropical*, **2014** 44(2), 175-181. doi: 10.1590/S1983-40632014000200005.
21. Botelho, C.E., Mendes A.N.G., Carvalho G.R., Bartholo G.F., & Carvalho S.P. Seleção de progênies F4 de cafeeiros obtidas pelo cruzamento de Icatu com Catimor. *Revista Ceres*, **2010a** 57(3), 274-281. doi: 10.1590/S0034-737X2010000300010
22. Carvalho, G.R., Botelho C.E., Bartholo G.F., Pereira A.A., Nogueira Â.M., & Carvalho, A.M. Comportamento de progênies F4 obtidas por cruzamentos de "Icatu" com "Catimor." *Ciência e Agrotecnologia*, **2009** 33(1), 47-52. doi: 10.1590/S1413-70542009000100006.
23. Pedro, F.C., Guimarães R.J., Carvalho G.R., Botelho C.E., de Rezende J.C., & Carvalho, A.M. Comportamento agrônomo de progênies F4 de cafeeiros oriundos do cruzamento entre os cultivares Mundo Novo e Catuaí. *Revista Ceres*, **2011** 58(3), 315-322. doi: 10.1590/S0034-737X2011000300011.
24. Rezende, R.M., Carvalho G.R., de Rezende J.C., Botelho C.E., Pasqualloto A.T., Velloso J.A., & Carvalho, A.M. de. Agronomic traits of coffee tree progenies from Timor Hybrid x Catuaí crossing. *Ciencia Rural*, **2014** 44(5), 775-780. doi: 10.1590/S0103-84782014000500003.
25. Raso, B.D.S.M., Botelho C.E., de Rezende J.C., Mendes A.N.G., Carvalho G.R., & Ferreira, A.D Avaliação agrônomo de cafeeiros da cultivar Mundo Novo no Estado de Minas Gerais. *Semina Ciencias Agrarias*, **2015** 36(4), 2363. doi: 10.5433/1679-0359.2015v36n4p2363.
26. Botelho, C.E., Rezende J.C. de, Carvalho G.R., Carvalho A.M. de, Andrade V.T., & Barbosa, C.R. Adaptabilidade e estabilidade fenotípica de cultivares de café arábica em Minas Gerais. *Pesquisa Agropecuária Brasileira*, **2010** 45(12), 1404-1411. doi: 10.1590/S0100-204X2010001200010.
27. Pereira, T.B., Mendes, A.N.G., Botelho, C.E., Rezende, J.C. de, Vilela, D.J.M., & Resende, M.D.V. Seleção de progênies F4 de cafeeiros obtidas de cultivares do grupo Icatu. *Coffee Science*, **2013** 8, 337-346.
28. Stern, A.J., Speidel L., Zaitlen N.A., & Nielsen R. Disentangling selection on genetically correlated polygenic traits via whole-genome genealogies. *The American Journal of Human Genetics*, **2021**, 108(2), 219-239. doi: 10.1016/j.ajhg.2020.12.005.