

Setting of Control Parameters of SOMA on the Base of Statistics

Zuzana Čičková and Martin Lukáčik

Abstract Evolutionary techniques are generally considered to be effective tool for solving a wide range of optimization problems. However, those algorithms are controlled by a special set of parameters according to their type. Control parameters of self-organizing migrating algorithm (SOMA) can be divided into several groups: the stopping parameters, parameters which depended on the type of problem to be solved and finally, parameters that are responsible for the quality of the results. The values of some parameters are directly evident from the nature of the algorithm, but the values of some may vary based on the problem and their efficient settings may significantly affect the quality of the calculation. This chapter focuses on the possibility of using some statistical methods to determine the effective values of some parameters of SOMA. The use of statistical methods is elucidated by an illustrative example.

1 Introduction

Evolutionary algorithms are successfully used for solving optimization problems of different types. Their limitation is caused by the fact that they are controlled by special set of parameters. Some of these parameters can be successfully set exogenously based on the philosophy of the algorithm, however, there is a no deeper theoretical base to adjust certain parameters (e.g. parameters determining the rate stochastics), whilst (im)proper setting can radically affect the quality of obtained results.

Z. Čičková (✉) · M. Lukáčik
Department of Operations Research and Econometrics, University of Economics
in Bratislava, Dolnozemska Cesta 1/B, 852 35 Bratislava, Slovakia
e-mail: cickova@euba.sk

M. Lukáčik
e-mail: martin.lukacik@euba.sk

Based on the various tests one can conclude that SOMA is even more sensitive to the parameters setting than other algorithms [3]. The control parameters are usually set on the basis of experimental results [3, 5]. Some of the control parameters are given directly by the nature of the problem and can be changed only by its reformulation. An example of such a parameter is the dimensionality (Dim). Setting other parameter can be derived from simple geometric interpretation of SOMA. Such parameter is the parameter PathLength, where its recommended setting is 3–5. Parameters PopSize and Migrations determine “the size and length” of simulation and their settings can use philosophy “more is better” (however, increasing these parameters affect the time needed to calculation and thus is dependent on the user’s hardware). The parameter MinDiv can be set to e.g. negative value if it is desired to reach all iterations, or to positive number if one want to watch the convergence of the calculation. Parameters Step and PRT are also responsible for the quality of the results. This chapter is devoted to some statistical methods that may be helpful in clarifying their settings. To adjust the control parameters it can be suitable before final calculating to carry out several simulations with e.g. smaller population size and lower number of iterations (which are not time consuming) with different values of the other control parameters. Further on, except basic descriptive statistics (e.g. average, mode, median), which allow to acquire the initial idea of the parameters settings, also various statistical methods can be used, e.g. single and multiple-factor analysis of variance.

The chapter is divided as follows. The first part is devoted to the theoretical description of some statistical methods. The second part gives an illustrative example of setting of control parameters.

2 Single and Multiple-Factor Analysis of Variance—Theory

Analysis of variance (ANOVA) is a technique, which enables to identify if there is any difference between groups on some variable (so called factor). When two or more groups are being compared, the characteristic that distinguishes the group from one another is called the factor under investigation. Consider the evolutionary techniques; an experiment might be carried out to compare different values of control parameters of algorithm from the perspective of obtained value of the fitness of the best individual (usually value of objective function).

Further on, the following notation will be used: a *population* is the set of all observations of interest and a sample is any subset of observations selected from the *population*. Let N be the total number of observation in the data set. Consider k levels of factor under investigation and a sample for each factor level, so that the sample size by j th factor level, $j = 1, 2, \dots, k$ is designate as n_j , $\sum_{j=1}^k n_j = N$. Then, the i th observation for each j th factor level can be designated as x_{ij} , $j = 1, 2, \dots, k$, $i = 1, 2, \dots, n_j$. Whether the null hypothesis of a single-factor analysis of variance

should be rejected depends on how substantially the samples from the different *populations* differ from one another. Let $\mu_j, j = 1, 2, \dots, k$ be a mean of *population* group on corresponding factor level.

A single-factor analysis of variance problem involves a comparison of all k group means. The objective is to test the null hypothesis (H_0):

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \quad (1)$$

against alternative hypothesis (H_a):

$$H_a: \text{at least two of the } \mu_j \text{'s, } j = 1, 2, \dots, k, \text{ are different} \quad (2)$$

A measure of disparity among the sample means is the between-group sum of squares, denoted by SSB and given by

$$SSB = \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 \quad (3)$$

where \bar{x}_j is the sample mean of j th group and $\bar{\bar{x}}$ is the overall mean (ratio of sum of all observations to the total number of observations in the data set). SSB has an associated degree of freedom ($df_1 = k - 1$).

A measure of variation within the k samples, called error sum of squares and denoted by SSE , is given by

$$SSE = \sum_{j=1}^k (n_j - 1) s_j^2 \quad (4)$$

where s_j^2 is the sample variance of j th group. SSE has an associated degree of freedom ($df_2 = N - k$).

Total sum of squares, denoted by SST , is given by

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2 \quad (5)$$

with associated degree of freedom ($df = N - 1$).

The relationship between those three sums of squares is called the fundamental identity and for single-factor analysis of variance is $SST = SSB + SSE$.

A mean square is a sum of squares divided by its degree of freedom. In particular:

$$\text{between-group mean square: } MSB = \frac{SSB}{k - 1}$$

$$\text{within-group mean square: } MSE = \frac{SSE}{N - k}$$

The test statistic (F) of the single-factor analysis of variance has a Fisher distribution and it is given by the formula: $F = \frac{MSB}{MSE}$.

The validity of the analysis of variance test requires some assumptions. Peck et al. [4] present these ones:

1. Each of the k group or *population* distributions is normal.
2. The k normal distributions have identical standard deviations.
3. The observations in the sample from any particular one of the k groups or *populations* are independent of one another.
4. When comparing group or *population* means, k random samples are selected independently of one another.

The statistical significance of the F ratio is most easily judged by its P -value. If the P -value is less than 0.05, the null hypothesis of equal means is rejected at the 5 % significance level. This does not imply that every mean is significantly different from every other mean. It only implies that the means are not all the same.

All the sums of squares, degrees of freedom, mean squares and F ratio with its P -value are entered in a general format of an analysis of variance table (Table 1).

Peck, Olsen and Devore also claim that in practice, the test based on these assumptions works well as long as the assumptions are not too badly violated. If the sample sizes are reasonably large, normal probability plots or boxplots of the data in each sample are helpful in checking the assumption of normality. Often, however, sample sizes are so small, that they suggest that the F test can safely be used if the largest of the sample standard deviations is at most twice the smallest one.

When null hypothesis is rejected by the F test, it can be stated that there are differences among the k group or *population* means. Several procedures called multiple-comparison procedures exist to determine which sample means are significantly different from others. Dowdy et al. [2] discuss five different approaches: Fisher's least significant difference, Duncan's new multiple-range test, Student–Newman–Keuls' procedure, Tukey's honestly significant difference and Scheffé's method.

Next, Fisher's least significant difference (LSD) procedure will be performed. Fisher's LSD procedure could be based on the t test statistic used for the *two-population* case. It could be easier to determine how large the difference between the sample means must be to reject null hypothesis.

In this case the test statistic by Anderson et al. [1] is the difference $\bar{x}_j - \bar{x}_l$, where $j, l = 1, 2, \dots, k$, so that $j \neq l$ and the objective is to test the null hypothesis (H_0):

Table 1 General format for an analysis of variance table

Source of variation	Sum of squares	Degree of freedom	Mean square	F -ratio	P -value
Between groups	SSB	$k - 1$	MSB	F	P
Within groups	SSE	$N - k$	MSE		
Total	SST	$N - 1$			

$$H_0: \mu_j = \mu_l; \quad j, l = 1, 2, \dots, k; \quad j \neq l \quad (6)$$

against alternative hypothesis (H_a):

$$H_0: \mu_j \neq \mu_l; \quad j, l = 1, 2, \dots, k; \quad j \neq l \quad (7)$$

The null hypothesis should be rejected if $|\bar{x}_j - \bar{x}_l| \geq LSD$, where least significant difference is given by

$$LSD = t_{\alpha/2} \sqrt{MSE \left(\frac{1}{n_j} + \frac{1}{n_l} \right)}; \quad j, l = 1, 2, \dots, k; \quad j \neq l \quad (8)$$

where α denotes the significance level and t denotes critical value of Student's distribution.

Dowdy, Weardon and Chilko recall, that Fisher's test has a drawback; it requires that the null hypothesis be rejected in the analysis of variance procedure by the F test. These authors also discuss presented assumptions. At first, the normality of the treatment groups can be roughly checked by constructing histograms of the sample from each group. The analysis of variance by them leads to valid conclusions in some cases where there are departures from normality. For small sample sizes the treatment groups should be symmetric and unimodal. For large samples, more radical departures are acceptable due to the central limit theorem. Dowdy, Weardon and Chilko assume that conditions on independence are usually satisfied if the experimental units are randomly chosen and randomly assigned to the treatments. If the treatment groups already exist the experimenter by them does not have the opportunity to assign the subjects at random to the treatments. In such cases he uses random samples from each treatment group.

The last one of the assumptions underlying the analysis of variance is that the variances of the *populations* from which the samples come are the same. Dowdy, Weardon and Chilko state that the F tests are robust with respect to departures from homogeneity; that is, moderate departures from equality of variances do not greatly affect the F statistic. If the experimenter fears a large departure from homogeneity, several procedures are available to test equality of variances.

The F_{max} test was developed by Hartley. Hartley's test may be used when all treatment groups are the same size n and involves comparing the largest sample variance with the smallest sample variance. The null hypothesis (H_0) of test (where σ_j^2 is the *population* variance of j th group, $j = 1, 2, \dots, k$) is:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \quad (9)$$

against alternative hypothesis (H_a):

$$H_a: \text{at least two of the } \sigma_j \text{'s, } j=1, 2, \dots, k, \text{ are different} \quad (10)$$

when each of the k populations is normal and there is a random sample of size n from each population. Then sample variances $s_j^2, j=1, 2, \dots, k$ can be computed and it is possible to calculate

$$F_{\max} = \frac{\max\{s_j^2, j=1, 2, \dots, k\}}{\min\{s_j^2, j=1, 2, \dots, k\}} \quad (11)$$

Statistics F_{\max} is significant if it exceeds the value given in the Fisher's table with degrees of freedom $df_1 = k$ and $df_2 = n - 1$. Dowdy, Weardon and Chilko state, that because of the sensitivity of Hartley's test to departures from normality, if statistics F_{\max} is significant, it indicates either unequal variances or a lack of normality.

Two other commonly used tests of homogeneity of variances are those of Cochran and Bartlett. In most situations, Cochran's test is equivalent to Hartley's. Cochran's test compares the maximum within-sample variance to the average within-sample variance. After computing of sample variances $s_j^2, j=1, 2, \dots, k$, it is calculated

$$C = \frac{\max\{s_j^2, j=1, 2, \dots, k\}}{\sum_{j=1}^k s_j^2} \quad (12)$$

and statistics C is significant if value

$$A = (k - 1) \frac{C}{1 - C} \quad (13)$$

exceeds the Fisher's table value with degrees of freedom $df_1 = \frac{n}{k} - 1$ and $df_2 = (\frac{n}{k} - 1)(k - 1)$.

Bartlett's test has a more complicated test statistic but has two advantages over the other two: It can be applied to groups of unequal sample sizes, and it is more powerful. Bartlett's test: compares a weighted average of the within-sample variances to their geometric mean. The test statistics is

$$B = \frac{1}{D} \left[\left(\sum_{j=1}^k (n_j - 1) \right) \ln \left(\frac{1}{\sum_{j=1}^k (n_j - 1)} \sum_{j=1}^k (n_j - 1) s_j^2 \right) - \sum_{j=1}^k (n_j - 1) \ln(s_j^2) \right] \quad (14)$$

where

$$D = 1 + \frac{1}{3(k-1)} \left[\sum_{j=1}^k \left(\frac{1}{n_j - 1} \right) - \frac{1}{\sum_{j=1}^k (n_j - 1)} \right] \tag{15}$$

Statistics B is significant if it exceeds the value given in the chi-squared distribution with $(k - 1)$ degrees of freedom.

The last presented statistics of homogeneity of variances is Levene’s test. This test performs a one-way analysis of variance on the variables $z_{ij} = |x_{ij} - \bar{x}_j|$, $j = 1, 2, \dots, k$, where \bar{x}_j is either a mean of j th group or a median of j th group. At first variables z_{ij} are computed and the F statistic of the single-factor analysis of variance for these variables is obtained. Levene’s statistics is significant if it exceeds the Fisher’s table value with degrees of freedom $df_1 = k - 1$ and $df_2 = N - k$.

Some statistical software also presents the results of a set of two-sample F tests that compare the standard deviations for each pair of levels. This makes sense only if the initial overall test shows significant differences amongst the variances (and standard deviations). Any pair with a small P value would be a pair whose standard deviations were significantly different.

An alternative to the standard analysis of variance that compares level medians instead of means is the Kruskal-Wallis test. This test is much less sensitive to the presence of outliers than a standard one-way analysis of variance and should be used whenever the assumption of normality within levels is not reasonable. Dowdy, Weardon and Chilko indicate this procedure.

First, it is necessary to rank the data from 1 (the smallest observation) to N (the largest observation), irrespective of the group in which they are found. If two or more observations are tied for the same numerical value, the average rank for which they are tied is assigned. Then for every group, when all treatment groups are the same size n , the average rank of group denoted by \bar{r}_j is computed. Finally, the test statistic is:

$$H = n \frac{\left[\sum_{j=1}^k (\bar{r}_j - \frac{N+1}{2})^2 \right]}{\frac{N(N+1)}{12}} \tag{16}$$

The null hypothesis (H_0) of test is

$$H_0: E(\bar{r}_j) = \frac{N+1}{2} \quad \text{for all } j \tag{17}$$

against alternative hypothesis (H_a):

$$H_a: E(\bar{r}_j) \neq \frac{N+1}{2} \quad \text{for some } j \tag{18}$$

where $E(\bar{r}_j)$ denotes the expected value by \bar{r}_j .

Statistics H is significant if it exceeds the value given in the chi-squared distribution with $(k - 1)$ degrees of freedom and it indicates that there are significant differences amongst the level medians.

However, in some experiments it is desirable to draw conclusions about more than one variable or factor. The term factorial is used because the experimental conditions include all possible combinations of the factors. For example, for a levels of factor A and b levels of factor B , the experiment will involve collecting data on ab combinations. In experimental design terminology, the sample size of r for each group combination indicates that there are r replications, so abr observations are needed. Additional replications ($2r$, $3r$) and larger sample size put statistical conclusions in more precise terms.

This situation brings new effect—interaction effect. If the interaction effect has a significant impact, it can be concluded that the effect of the type of factor A depends on the factor B . There are three sets of hypothesis with the two-way ANOVA.

At first, the objective is to test the null hypothesis of comparison of a group means μ_{Ai} , $i = 1, 2, \dots, a$ by different values of factor A (H_0):

$$H_0: \mu_{A1} = \mu_{A2} = \dots = \mu_{Aa} \quad (19)$$

against alternative hypothesis (H_a):

$$H_a: \text{at least two of the } \mu_{Aj} \text{'s, } j = 1, 2, \dots, a, \text{ are different} \quad (20)$$

and also comparison of b group means μ_{Bi} , $j = 1, 2, \dots, b$ by different values of factor B :

$$H_0: \mu_{B1} = \mu_{B2} = \dots = \mu_{Bb} \quad (21)$$

against alternative hypothesis (H_a):

$$H_a: \text{at least two of the } \mu_{Bj} \text{'s, } j = 1, 2, \dots, b, \text{ are different} \quad (22)$$

Second objective is comparison of ab group means μ_{AiBj} , $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$ by different values of A and B :

$$H_0: \mu_{A1B1} = \mu_{A2B2} = \dots = \mu_{AaBb} \quad (23)$$

against the alternative (H_a):

$$H_a: \text{There is no interaction between the factors } A \text{ and } B \quad (24)$$

The analysis of variance procedure for the two-factor factorial experiment requires us to partition the total sum of squares into sum of squares for factor A , sum of squares for factor B , sum of squares for interaction and sum of squares due to error.

Table 2 General format for an analysis of variance table

Source of variation	Sum of squares	Degree of freedom	Mean square	F-ratio	P-value
Factor A	SSA	$a - 1$	MSA	F_A	P_A
Factor B	SSB	$b - 1$	MSB	F_B	P_B
Interaction	SSAB	$(a - 1)(b - 1)$	$MSAB$	F_{AB}	P_{AB}
Error	SSE	$ab(r - 1)$	MSE		
Total	SST	$N - 1$			

Sum of squares for factor A is denoted by *SSA* and given by

$$SSA = br \sum_{i=1}^a (\bar{x}_i - \bar{\bar{x}})^2 \tag{25}$$

where *b* is number of levels of factor B, *a* is number of levels of factor A, $\bar{\bar{x}}$ is the overall mean and *r* is the number of replications.

Sum of squares for factor B is denoted by *SSB* and given by

$$SSB = ar \sum_{i=1}^b (\bar{x}_i - \bar{\bar{x}})^2 \tag{26}$$

Sum of squares for interaction is denoted by *SSAB* and given by

$$SSAB = r \sum_{i=1}^a \sum_{j=1}^b (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2 \tag{27}$$

where x_{ij} is the sample mean for the observations corresponding to the combination of group *i* (factor A) and group *j* (factor B).

Error sum of squares (*SSE*) and total sum of squares (*SST*) are given by the same relations as in the case of single-factor analysis of variance (4) and (5). All the sums of squares, degrees of freedom, mean squares and *F* ratios with their *P*-values are presented in the analysis of variance table (Table 2).

3 Parameters Setting of SOMA

Next, the setting of control parameters of SOMA will be presented based on an illustrative example of solving traveling salesman problem. Consider the matrix of shortest distances between eight cities (Table 3).

The traveling salesman needs to find the shortest route between all the cities so that each city is visited exactly once. The solving was provided with the use of natural representation (the city was represented directly with its index in an

Table 3 Matrix of shortest distances between eight cities

	City 1	City 2	City 3	City 4	City 5	City 6	City 7	City 8
City 1	–	202	197	115	191	123	161	86
City 2	202	–	389	85	393	121	48	195
City 3	197	389	–	304	35	317	351	241
City 4	115	85	304	–	308	84	46	138
City 5	191	393	35	308	–	297	355	221
City 6	123	121	317	84	297	–	73	74
City 7	161	48	351	46	355	73	–	147
City 8	86	195	241	138	221	74	147	–

individual). A simple penalty approach was used if the unfeasible solutions appeared. Some of the control and termination parameters were set as follows: Parameter PopSize was set to 80 and parameter Migrations was set to 300. The parameter MinDiv was set to negative value to reach all iterations (the small size of instance enables to reach all iteration in a relative short time). The parameter PathLength was set to the value 3. The settings of parameters Step and PRT was provided on the base of before mentioned statistical methods. Let the value of the shortest route (denoted as fc) be the response variable. Further on, we can specify the impact of factors' level (levels of parameters Step and PRT) on the variability of the response variable.

Parameter PRT can take values from 0 (purely stochastic behavior of algorithm) to 1 (purely deterministic behavior). At first, the levels of parameters PRT were set to 0.2, 0.4, 0.6 and 0.8. Parameter Step can take values from 0.1 to value of parameter PathLength, which equals 3. Following the previous simulations it was found that the value of the Step >1, increased probability of getting extremely “bad” outcome. So, the values 0.3, 0.5, 0.7 and 0.9 were used as the levels of parameter Step in testing.

First tested hypothesis is comparison of 4 group means $\bar{fc}_1, \bar{fc}_2, \bar{fc}_3, \bar{fc}_4$ by different values of parameter PRT (0.2, 0.4, 0.6 and 0.8) according to (1) and (2). The experiment is balancing—for each pair PRT-Step the same number of simulations was realized—eight replications. It was thus implemented the first phase of a total of 128 simulations (Table 4).

The summary of the descriptive statistics for every value of parameter PRT is given in Table 5.

There is big difference between the smallest and the largest standard deviation. This may cause problems since the analysis of variance assumes that the standard deviations at all levels are equal. The results are also presented by the box and whisker plot (Fig. 1).

There is evident some significant non-normality in the data, which violates the assumption that the data come from normal distributions. Someone may wish to transform the values of fc to remove any dependence of the standard deviation on

Table 4 Results of simulations

Step	PRT	f_c	Step	PRT	f_c	Step	PRT	f_c	Step	PRT	f_c
0.3	0.2	862	0.3	0.2	862	0.3	0.2	921	0.3	0.2	906
0.3	0.4	862	0.3	0.4	848	0.3	0.4	848	0.3	0.4	848
0.3	0.6	848	0.3	0.6	848	0.3	0.6	848	0.3	0.6	848
0.3	0.8	848	0.3	0.8	848	0.3	0.8	848	0.3	0.8	848
0.5	0.2	896	0.5	0.2	857	0.5	0.2	1072	0.5	0.2	896
0.5	0.4	906	0.5	0.4	896	0.5	0.4	842	0.5	0.4	848
0.5	0.6	848	0.5	0.6	848	0.5	0.6	848	0.5	0.6	848
0.5	0.8	848	0.5	0.8	848	0.5	0.8	848	0.5	0.8	848
0.7	0.2	848	0.7	0.2	907	0.7	0.2	979	0.7	0.2	963
0.7	0.4	848	0.7	0.4	871	0.7	0.4	848	0.7	0.4	857
0.7	0.6	848	0.7	0.6	905	0.7	0.6	848	0.7	0.6	848
0.7	0.8	857	0.7	0.8	848	0.7	0.8	848	0.7	0.8	848
0.9	0.2	980	0.9	0.2	848	0.9	0.2	857	0.9	0.2	905
0.9	0.4	848	0.9	0.4	857	0.9	0.4	848	0.9	0.4	848
0.9	0.6	848	0.9	0.6	848	0.9	0.6	848	0.9	0.6	848
0.9	0.8	848	0.9	0.8	857	0.9	0.8	857	0.9	0.8	848
0.3	0.2	919	0.3	0.2	988	0.3	0.2	871	0.3	0.2	954
0.3	0.4	848	0.3	0.4	848	0.3	0.4	857	0.3	0.4	848
0.3	0.6	848	0.3	0.6	848	0.3	0.6	848	0.3	0.6	848
0.3	0.8	848	0.3	0.8	848	0.3	0.8	848	0.3	0.8	848
0.5	0.2	905	0.5	0.2	919	0.5	0.2	857	0.5	0.2	896
0.5	0.4	848	0.5	0.4	848	0.5	0.4	857	0.5	0.4	848
0.5	0.6	848	0.5	0.6	848	0.5	0.6	848	0.5	0.6	848
0.5	0.8	848	0.5	0.8	848	0.5	0.8	848	0.5	0.8	848
0.7	0.2	907	0.7	0.2	919	0.7	0.2	848	0.7	0.2	862
0.7	0.4	848	0.7	0.4	848	0.7	0.4	862	0.7	0.4	848
0.7	0.6	848	0.7	0.6	857	0.7	0.6	857	0.7	0.6	857
0.7	0.8	862	0.7	0.8	848	0.7	0.8	848	0.7	0.8	848
0.9	0.2	862	0.9	0.2	907	0.9	0.2	919	0.9	0.2	1003
0.9	0.4	848	0.9	0.4	848	0.9	0.4	848	0.9	0.4	905
0.9	0.6	848	0.9	0.6	848	0.9	0.6	848	0.9	0.6	848
0.9	0.8	848	0.9	0.8	848	0.9	0.8	848	0.9	0.8	857

the mean. The analysis of variance decomposes the variance of f_c into two components: a between-group component and a within-group component (Table 6).

The F ratio, which equals 33.0788, is a ratio of the between-group estimate to the within-group estimate. Since the P -value of the F test is less than 0.05, there is a statistically significant difference between the mean f_c from one level of PRT to another at the 5.0 % significance level. Then, Fisher’s least significant difference was used to determine which means are significantly different from which others (Table 7).

Table 5 Summary statistics—data grouped by PRT

PRT	Count	Average	Standard deviation	Coefficient of variation (%)	Minimum	Maximum	Range	Median
0.2	32	909.219	53.0324	5.83274	848.0	1072.0	224.0	905.5
0.4	32	855.625	16.3997	1.9167	842.0	906.0	64.0	848.0
0.6	32	850.625	10.2729	1.20769	848.0	905.0	57.0	848.0
0.8	32	849.563	3.77545	0.444399	848.0	862.0	14.0	848.0
Total	128	866.258	37.5064	4.3297	842.0	1072.0	230.0	848.0

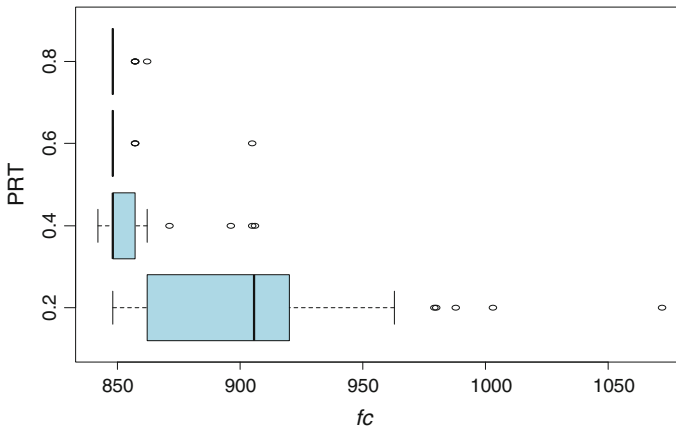


Fig. 1 Box and whisker plot—data grouped by PRT

Table 6 Analysis of variance table—data grouped by PRT

Source of variation	Sum of squares	Degree of freedom	Mean square	F-ratio	P-value
Between groups	79418.1	3	26472.7	33.08	0.0000
Within groups	99236.3	124	800.293		
Total	178654.4	127			

Table 7 Comparison procedure of Fisher’s least significant difference—data grouped by PRT

Contrast	Difference of means	LSD	Significant differences
0.2–0.4	53.5938	13.9982	Yes
0.2–0.6	58.5938	13.9982	Yes
0.2–0.8	59.6563	13.9982	Yes
0.4–0.6	5.0	13.9982	No
0.4–0.8	6.0625	13.9982	No
0.6–0.8	1.0625	13.9982	No

Now, one can see significant difference for group of simulations where PRT equals 0.2 to other groups. It can be stated that there is a large departure from homogeneity, so all the equality of variances' tests are used (Table 8).

The statistics displayed in the Table 8 and also the *P*-values show, that there is a statistically significant difference amongst the standard deviations of groups. This violates one of the important assumptions underlying the analysis of variance.

The comparison of the standard deviations for each pair of samples is given in Table 9. All *P*-Values below 0.05 indicate statistically significant differences between standard deviations of every pair of groups.

The situation is clear; statistically different values of averages and standard deviations for groups of values *f_c* by different values of parameter PRT were obtained. Due to failure of assumptions, the results of the analysis of variance cannot be taken into account. Finally, despite all the previous conclusions, the decision is to use the Kruskal-Wallis test as alternative to the standard analysis of variance to compare the medians instead of the means (Table 10).

Table 8 Tests of homogeneity of variances—data grouped by PRT

	Test	<i>P</i> -value
Levene's	23.1283	5.87053E-12
Cochran's	0.878564	0
Bartlett's	4.35742	0
Hartley's	197.308	

Table 9 Comparison of the standard deviations for each pair of groups—data grouped by PRT

Comparison	Standard deviation 1	Standard deviation 2	<i>F</i> -ratio	<i>P</i> -value
0.2/0.4	53.0324	16.3997	10.457	0.0000
0.2/0.6	53.0324	10.2729	26.65	0.0000
0.2/0.8	53.0324	3.77545	197.308	0.0000
0.4/0.6	16.3997	10.2729	2.54853	0.0111
0.4/0.8	16.3997	3.77545	18.8685	0.0000
0.6/0.8	10.2729	3.77545	7.40368	0.0000

Table 10 Kruskal-Wallis test —data grouped by PRT

PRT	Sample size	Average rank
0.2	32	104.469
0.4	32	57.4375
0.6	32	47.5
0.8	32	48.5938
Test statistic	66.7	<i>P</i> -value = 0

The null hypothesis of Kruskal-Wallis test is that the medians of fc within each of the four levels of PRT are the same (17). Since the P -value is less than 0.05, there is a statistically significant difference amongst the medians.

It is evident (from the results of tests and also from box and whisker plot) that median of group where PRT equals 0.2 is significantly different from others. It seems that values 0.6 or 0.8 for the parameter PRT are the appropriate choice. More preferred alternative is a latter value in order to eliminate possible outliers.

Second tested hypothesis is comparison of 4 group means $\bar{fc}_1, \bar{fc}_2, \bar{fc}_3, \bar{fc}_4$ by different values of parameter Step (0.3, 0.5, 0.7 and 0.9) according to (1) and (2).

The summary of the descriptive statistics by every value of Step is given in the Table 11.

In this case there is not so big difference between the smallest and the largest standard deviation as in previous case. From the box and whisker plot (Fig. 2) it is seen some significant non-normality in the data, which again violates the assumption that the data come from normal distributions. Recall that the normal distribution is symmetric with a median in the middle of the box bounded by the

Table 11 Summary statistics—data grouped by Step

Step	Count	Average	Standard dev.	Coef. of variation (%)	Minimum	Maximum	Range	Median
0.3	32	864.313	34.6656	4.01077	848.0	988.0	140.0	848.0
0.5	32	867.469	43.6984	5.03746	842.0	1072.0	230.0	848.0
0.7	32	866.813	33.9995	3.92236	848.0	979.0	131.0	848.0
0.9	32	866.438	38.6013	4.45517	848.0	1003.0	155.0	848.0
Total	128	866.258	37.5064	4.3297	842.0	1072.0	230.0	848.0

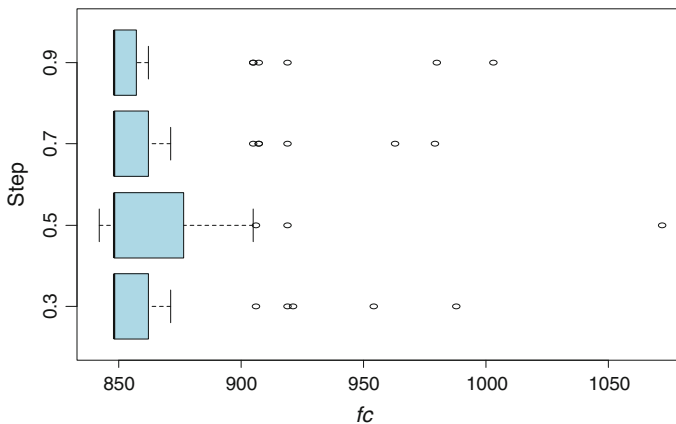


Fig. 2 Box and whisker plot—data grouped by Step

first and the third quartile. This is not the case, because the median is the smallest value in three cases and for the last one is typical outlier.

The analysis of variance decomposes the variance of fc once again into two components: a between-group component and a within-group component, but now data are grouped by parameter Step (Table 12).

The F ratio, which equals 0.0414313, is a ratio of the between-group estimate to the within-group estimate. Since the P -value of the F test is greater than 0.05, there is not a statistically significant difference amongst the mean fc from one level of Step to another at the 5.0 % significance level. Fisher’s test requires that the null hypothesis could be rejected in the analysis of variance procedure by the F test, what is not the case; nevertheless its results are shown (Table 13).

Evidently, there is not a significant difference in means between groups.

The all statistics displayed in the Table 14 and also the P -values greater than or equal to 0.05 show, that there is not a statistically significant difference amongst the standard deviations of groups.

The comparison of the standard deviations for each pair of samples is given in Table 15. It can be stated there are no statistically significant differences between any pair of means.

Table 12 Analysis of variance table—data grouped by Step

Source of variation	Sum of squares	Degree of freedom	Mean square	F -ratio	P -value
Between groups	178.898	3	59.6328	0.04	0.9887
Within groups	178476.6	124	1439.32		
Total	178654.4	127			

Table 13 Comparison procedure of Fisher’s least significant difference—data grouped by Step

Contrast	Difference of means	LSD	Significant differences
0.3–0.5	-3.15625	18.7727	No
0.3–0.7	-2.5	18.7727	No
0.3–0.9	-2.125	18.7727	No
0.5–0.7	0.65625	18.7727	No
0.5–0.9	1.03125	18.7727	No
0.7–0.9	0.375	18.7727	No

Table 14 Tests of homogeneity of variances—data grouped by Step

	Test	P -value
Levene’s	0.0490732	0.985551
Cochran’s C	0.331675	0.298469
Bartlett’s	1.02079	0.472019
Hartley’s	1.65191	

Table 15 Comparison of the standard deviations for each pair of groups—data grouped by Step

Comparison	Standard deviation 1	Standard deviation 2	F-ratio	P-value
0.3/0.5	34.6656	43.6984	0.629314	0.2029
0.3/0.7	34.6656	33.9995	1.03957	0.9147
0.3/0.9	34.6656	38.6013	0.806481	0.5529
0.5/0.7	43.6984	33.9995	1.65191	0.1679
0.5/0.9	43.6984	38.6013	1.28152	0.4939
0.7/0.9	33.9995	38.6013	0.775783	0.4838

The situation is different from the previous case; we didn't obtain statistically different values of averages and standard deviations for groups of values fc by different values of parameter Step. Again, we decided to use alternative to the standard analysis of variance—the Kruskal-Wallis test—to compare the medians instead of the means (Table 16).

The null hypothesis of Kruskal-Wallis test is that the medians of fc within each of the four levels of Step are the same (17). Since the P -value is greater than 0.05, there is not a statistically significant difference amongst the medians.

Hence, the results of tests and also from box and whisker plot show the means, medians and standard deviations of all four samples are equal. It is not a difference between arbitrary values of the parameter Step from a statistical point of view. Despite the results, it seems that values 0.7 or 0.9 for the parameter Step is the appropriate choice, since calculations are usually faster for bigger values of Step. These values generate the equivalent results with similar outliers. More preferred alternative is a latter value because of smaller interquartile range.

Last tested hypothesis is comparison of 4 group means $\bar{fc}_{A1}, \bar{fc}_{A2}, \bar{fc}_{A3}, \bar{fc}_{A4}$ by different values of parameter Step (factor A) according the test (19) and (20), where the levels of Step were set to 0.3, 0.5, 0.7 and 0.9 and also comparison of 4 group means $\bar{fc}_{B1}, \bar{fc}_{B2}, \bar{fc}_{B3}, \bar{fc}_{B4}$ by different values of parameter PRT (factor B) according the test (21) and (22), where the levels of PRT were set to 0.2, 0.4, 0.6 and 0.8, as well as the comparison of 16 group means $\bar{fc}_{A1B1}, \bar{fc}_{A2B1}, \dots, \bar{fc}_{A4B4}$ by mentioned different values of Step and PRT (23) and (24).

The ANOVA table (Table 17) decomposes the variability of fc into contributions due to both factors Step and PRT. The contribution of each factor is measured having removed the effect of another factor. Since P -value of factor PRT is less than 0.05, this factor has a statistically significant effect on fc at the 5.0 % significance

Table 16 Kruskal-Wallis test —data grouped by Step

Step	Sample size	Average rank
0.3	32	62.0625
0.5	32	61.6875
0.7	32	69.875
0.9	32	64.375
Test statistic	1.30262	P -value = 0.728508

Table 17 Multiple analysis of variance table

Source	Sum of squares	Df	Mean square	F-ratio	P-value
Factor A: step	178.898	3	59.6328	0.07	0.9766
Factor B: prt	79418.1	3	26472.7	30.35	0.0000
Interactions AB	1359.82	9	151.091	0.17	0.9964
Error	97697.6	112	872.3		
Total	178654	127			

level. Significant interaction effects between analysed factors have not been confirmed. The results of multiple factor analysis of variance confirmed the conclusions that were obtained using a single factor analysis of variance. Different values of parameter Step didn't result to statistically different values of function fc . Contrary, different values of parameter PRT resulted to statistically different values of function fc .

It is evident (Fig. 3) that the small values of PRT (0.2 and 0.4) results to big variability of values fc regardless of the value of Step. The interaction plot (Fig. 4) gives the mean values of fc depending on combination of mentioned factors. Based on that, it seems to be an appropriate choice to set the parameter PRT to 0.6 or 0.8.

Further on, one more analysis was realized in order to choose between the two values of parameter PRT and one way factor analysis for parameter PRT was conducted. Tested hypothesis is comparison of 5 group means $\bar{fc}_1, \bar{fc}_2, \bar{fc}_3, \bar{fc}_4, \bar{fc}_5$ by different values of parameter PRT on levels 0.5, 0.6, 0.7, 0.8 and 0.9 according to (1) and (2). Parameter Step was set to value 0.9. The experiment is balancing—for each value of PRT the same number of simulations was realized—eight replications. It was thus implemented a total of 40 simulations, which results are summarized in Table 18.

The summary of the descriptive statistics by every value of parameter PRT can be seen in the Table 19.

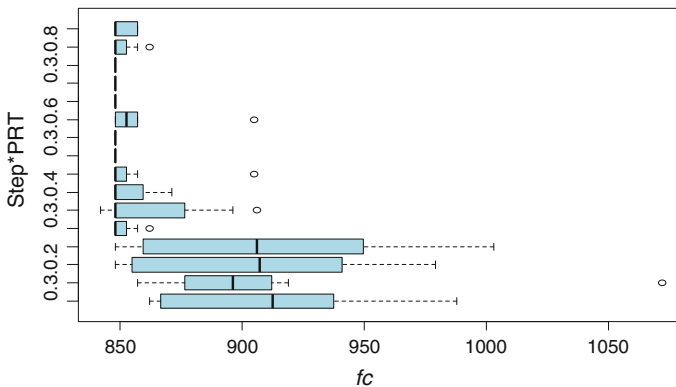


Fig. 3 Box and whisker plot—data grouped by interaction of Step and PRT

Fig. 4 Interaction plot

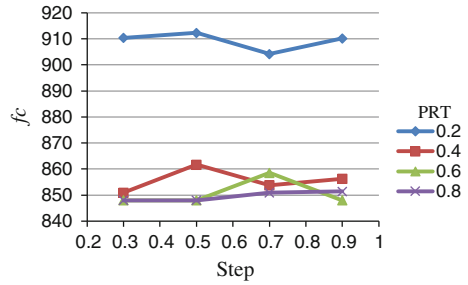


Table 18 Results of simulations

PRT	<i>fc</i>	PRT	<i>fc</i>	PRT	<i>fc</i>	PRT	<i>fc</i>	PRT	<i>fc</i>
0.5	848	0.6	857	0.7	848	0.8	848	0.9	848
0.5	857	0.6	848	0.7	848	0.8	848	0.9	848
0.5	862	0.6	921	0.7	857	0.8	857	0.9	848
0.5	862	0.6	848	0.7	848	0.8	848	0.9	848
0.5	959	0.6	896	0.7	857	0.8	848	0.9	848
0.5	905	0.6	871	0.7	848	0.8	857	0.9	848
0.5	1007	0.6	857	0.7	848	0.8	857	0.9	848
0.5	848	0.6	848	0.7	857	0.8	848	0.9	848

Table 19 Summary statistics—data grouped by PRT

prt	Count	Average	Standard dev.	Coef. of variation (%)	Minimum	Maximum	Range	Median
0.5	8	893.5	59.4763	6.65655	848.0	1007.0	159.0	862.5
0.6	8	868.25	26.8421	3.09152	848.0	921.0	73.0	857.0
0.7	8	851.375	4.65794	0.547108	848.0	857.0	9.0	848.0
0.8	8	851.375	4.65794	0.547108	848.0	857.0	9.0	848.0
0.9	8	848.0	0	0	848.0	848.0	0	848.0
Total	40	862.5	32.7085	3.79229	848.0	1007.0	159.0	848.0

There is again big difference between the smallest and the largest standard deviation. Remember, that this may cause problems since the analysis of variance assumes that the standard deviations at all levels are equal. It is evident also from the box and whisker plot of results (Fig. 5).

It is evident there is some significant non-normality in the data, which violates the assumption that the data come from normal distributions. The analysis of variance decomposes the variance of *fc* into two components: a between-group component and a within-group component (Table 20).

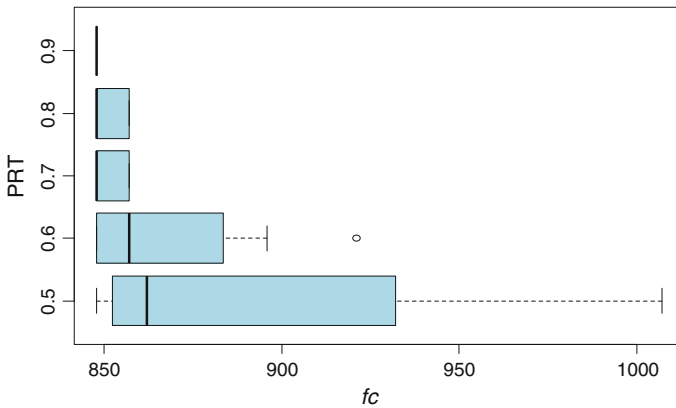


Fig. 5 Box and whisker plot—data grouped by PRT

Table 20 Analysis of variance table—data grouped by PRT

Source of variation	Sum of squares	Degree of freedom	Mean square	<i>F</i> -ratio	<i>P</i> -value
Between groups	11614.8	4	2903.69	3.3753	0.01946
Within groups	30109.3	35	860.264		
Total	41724.0	39			

The *F* ratio, which equals 3.3753, is a ratio of the between-group estimate to the within-group estimate. Since the *P*-value of the *F* test is less than 0.05, there is a statistically significant difference between the mean *fc* from one level of PRT to another at the 5.0 % significance level. Fisher’s least significant difference is used to determine which means are significantly different from which others (Table 21).

It is seen significant difference for group of simulations where PRT equals 0.5 to groups where PRT equal to 0.7, 0.8 and 0.9. It is evident a large departure from homogeneity, so next all the equality of variances’ tests are used (Table 22).

The statistics displayed in this table and also the *P*-values show, that there is a statistically significant difference amongst the standard deviations of groups. This violates one of the important assumptions underlying the analysis of variance.

The comparison of the standard deviations for each pair of samples is given in Table 23. *P*-Values below 0.05 indicate statistically significant differences between standard deviations of these pair of groups.

The situation is such as in the first analysis of parameter PRT; there are statistically different values of averages and standard deviations for groups of values *fc* by different values of parameter PRT. Due to failure of assumptions, the results of the analysis of variance cannot be taken into account. Finally, despite all the previous conclusions, we decided to use the Kruskal-Wallis test to compare the medians instead of the means (Table 24).

Table 21 Comparison procedure of Fisher's least significant difference—data grouped by PRT

Contrast	Difference of means	LSD	Significant differences
0.5–0.6	25.25	29.7719	No
0.5–0.7	42.125	29.7719	Yes
0.5–0.8	42.125	29.7719	Yes
0.5–0.9	45.5	29.7719	Yes
0.6–0.7	16.875	29.7719	No
0.6–0.8	16.875	29.7719	No
0.6–0.9	20.25	29.7719	No
0.7–0.8	0	29.7719	No
0.7–0.9	3.375	29.7719	No
0.8–0.9	3.375	29.7719	No

Table 22 Tests of homogeneity of variances—data grouped by PRT

	Test	P-value
Levene's	3.2163	0.023823
Cochran's	0.822405	0.00000532605
Bartlett's	5.77786	4.76942E-10
Hartley's	163.042	

Table 23 Comparison of the standard deviations for each pair of groups – data grouped by PRT

Comparison	Standard deviation 1	Standard deviation 2	F-ratio	P-value
0.5/0.6	59.4763	26.8421	4.90969	0.0523
0.5/0.7	59.4763	4.65794	163.042	0.0000
0.5/0.8	59.4763	4.65794	163.042	0.0000
0.5/0.9	59.4763	0	–	–
0.6/0.7	26.8421	4.65794	33.2082	0.0002
0.6/0.8	26.8421	4.65794	33.2082	0.0002
0.6/0.9	26.8421	0	–	–
0.7/0.8	4.65794	4.65794	1.0	1.0000
0.7/0.9	4.65794	0	–	–
0.8/0.9	4.65794	0	–	–

The null hypothesis of Kruskal-Wallis test (17) is that the medians of fc within each of the five levels of PRT are the same. Since the P -value is less than 0.05, there is a statistically significant difference amongst the medians. It seems that median of group where PRT equals 0.5 and 0.6 are significantly different from others. Based on mentioned above the values 0.7–0.9 for the parameter PRT are considered as the appropriate choice.

Table 24 Kruskal-Wallis test
—data grouped by PRT

PRT	Sample size	Average rank
0.5	8	29.375
0.6	8	25.125
0.7	8	18.0
0.8	8	18.0
0.9	8	12.0
Test statistic	10.844	P -value = 0.0283745

4 Conclusions

Evolutionary algorithms are considered to be universal and effective tool for solving various optimization problems. Their effectiveness is limited by fact they are generally controlled by special set of parameters. Although some of parameters can be successfully set exogenously based on the philosophy of the algorithm or according to type of solved problem, there is a no deeper theoretical base to adjust all the parameters. This chapter focuses on the possibility of using some statistical methods that may be helpful to determine the effective values of some parameters of SOMA.

Based on the various tests one can conclude that SOMA is even more sensitive on the parameters setting than other algorithms [3, 5], thus the efficient setting may significantly affect the quality of the results. The setting of control parameter can be supported by statistical methods especially aimed at determining whether the level of some parameter brings the difference in results. A brief view to corresponding statistical methods (single factor analyze of variance, Levene's test, Cochran's test, Bartlett's test, Hartley's test, two-way analyze of variance) is given in the first half of the chapter. The second half is aimed on example of practical use based on illustrative data of traveling salesman problem.

References

1. Anderson, D.R., Sweeney, D.J., Williams, T.A.: Statistics for business and economics, 11th edn. South-Western Cengage Learning, Boston (2011)
2. Dowdy, S., Weardon, S., Chilko, D.: Statistics for research, 3rd edn. Wiley, New York (2004)
3. Onwubolu, G.C., Babu, B.V.: New Optimization Techniques in Engineering. Springer, Berlin (2004)
4. Peck, R., Olsen, C., Devore, J.: Introduction to statistics and data analysis, 4th edn. Cengage Learning, Boston (2012)
5. Zelinka, I.: Umělá inteligence v problémech globální optimalizace. BEN-technická literature, Praha (2002)