MONTE CARLO SIMULATION FOR ANOVA
Gabriela Ižaríková

MONTE CARLO SIMULATION FOR ANOVA

Gabriela Ižaríková

TU of Košice, Faculty SjF, Institute of Special Technical Sciences, Department of Applied Mathematics and Informatics, Letná 9, 042 00 Košice, gabriela.izarikova@tuke.sk

Keywords: One-way ANOVA, simulation, Monte Carlo method
Abstract: The article we will present Monte Carlo simulation for assessing consequences of data non assumption. Analysis of variance (ANOVA)) is used to determine whether there are any significant differences between the means of three or more independent (unrelated) groups. Fundamental assumption for ANOVA is that the independent variable is normaly distributed and groups with equal variances. Monte Carlo simulation we observed Type I. error rate of analysis of variance.

1 Introduction

Using quantitative mathematical or statistical model looking for the best optimal solution. In practice it is often unrealistic to find an optimal solution, for example, there are used the basic conditions of the methods. It is possible to use simulating models to solve these situations. Simulating real problems is one of the most frequently used approach that facilitates decision-making. The simulation model generally shows a system modeled using mathematical formulations and logical relationships. In the model we distinguish between a random input to control, that is transformed to output model. For the simulation experiment in the beginning you select controlled access and random (stochastic) inputs are randomly generated.

Simulations are among the quantitative tools that can be used for decision support. Simulation of work with a particular model is an experiment with the model. It is a subset of simulation modeling allows broaden the scope of the investigation and of the specific model types. Monte Carlo simulation method is known that uses a large number of randomly generated samples from the probability distribution that is used for computer simulation solutions to various managerial problems from mathematics, physics, financing, design, sales, human resources, psychology and other [1], [2].

In statistical theory, we meet two basic types of methods: parametric and nonparametric. Parametric methods (tests) are characterized in that they comply with certain assumptions. If you fulfill the requirements of the methods, such as processed data come from a normal distribution, statistical methods offer an effective and valid estimates of the probability distribution of statistics [3]. When the theoretical assumptions do not examine data, then the validity of the statistics reliable estimates of the probability distribution is uncertain. In such situations, it is possible to use Monte Carlo simulations [4], [5]. This method favors empirical estimates statistics probability distribution file prior theoretical expectations on these figures. The essence of Monte Carlo method is that it generates numerous scenarios studied a random file. Using Monte Carlo simulation can demonstrate how to approach the theoretical results. In this paper, the Monte Carlo method applied to a situation where the assumptions are not met statistical methods, namely analysis of variance (ANOVA).

Monte Carlo method comprises the following steps:
- ➢ Determine the objective of the simulation.
- ➢ Propose appropriate methods of Monte Carlo.
- ➢ On the basis of concrete statistics randomly generated data.
- ➢ Implement quantitative methods.
- ➢ Quantify the necessary statistics.
- ➢ Simulation contrary (eg. 100 to 1 000,000 times).
- ➢ Analysis of statistics found.
- ➢ Assess the results obtained by the methods of Monte Carlo.

2 One-way ANOVA

The One-way ANOVA (analysis of variance) is used to determine whether there are any significant differences between the means of three or more independent (unrelated) groups. The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are significantly different from each other [6]. Specifically, it tests the null hypothesis (1):

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots \mu_k \qquad (1)$$
$$\text{and then} \quad H_1: \text{non } H_0$$

where μ - group mean and k - number of groups. If, however, the one-way ANOVA returns a significant result, we accept the alternative hypothesis (H_1), which is that there are at least two group means that are significantly different from each other. Its aim is to detect whether any differences between the means for these files are statistically significant or only incidental. Analysis of variance was trying to figure out which of quantitative or

MONTE CARLO SIMULATION FOR ANOVA
Gabriela Ižaríková

qualitative factors significantly influence the monitored variables.

The basic assumptions for the use of analysis of variance include:

➤ Independence of observations - the individual selections are independent of each other.
➤ Normality of sampling distribution - the samples come from a core set of normal distribution.
➤ Homogeneity of variance (homoscedasticity) – equal variances.

Number of factors examined by analysis of variance divided on:

➤ One-way analysis of variance - if the observed effect of one factor.
➤ Multi-factorial analysis of variance - for the monitoring of the impact of several factors.

According to distinguish the range of sample:

➤ Balanced model - if the coverage is the same sample.
➤ Unbalanced model- if different range of sample.

Random selection of independence is considered logically and ensure the appropriate selection of files. To verify in practice the second and third condition. Whether the results are valid in ANOVA failing these checked by using the Monte Carlo simulation, in which will be monitored type I error by One-way ANOVA (p-value).

2.1 *Various alternative for simulation by Monte Carlo methods*

Specifically, in this paper it is to test the hypothesis of conformity means of three groups:

$$H_0: \mu_1 = \mu_2 = \mu_3 = 100$$

Compared to the alternative hypothesis that the at least two diameters are equal. The simulation method Monte Carlo we will consider all alternatives that may arise, this means meeting respectively. failure to comply with terms and conditions of normality data homogeneity of each group.

Consider the following alternatives: (Table 1)

• The data come from a normal distribution with means (averagea) $\mu = 10$.
• The data come from the division that has skewness $\gamma_3 = 1.15$ and kurtosis $\gamma_4 = 2$.
• Where equal variances we suppose
$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 25.$$
• In case of different variances we suppose
$$\sigma_1^2 = 4, \sigma_2^2 = 25, \sigma_3^2 = 49.$$
• Assume that the individual files have the same number of observations, for example, the twenty.

Probability density graphs for all possible alternatives are on the Figure 1. Normal distribution is bell-shaped, which takes a maximum at x=μ. The hill is steepened when variances are smaller.

The assumptions of normality can be tested e.g. Shapiro-Wilk test. The Shapiro–Wilk test utilizes the null hypothesis principle to check whether a sample x_1, ..., x_n came from a normally distributed population. Result of test is *p-value*, if p <α (α = 0,05) and the null hypothesis is rejected on the significance level is 0.05. Test results for the various alternatives are in Table 2. Alternatives C and D does not satisfy the condition of normality.

The equal of variances basic set can be determined by Bartlett's test, it's a universal test that can be used for assessing the homogeneity of variances, but is relatively weak and quite sensitive to the violation of normality files, which can be a problem for files with a small number of observations. If the frequency of all choose the same used to test Cochrane test or Hartley test. The most commonly used test for homogeneity of variance test is Leveneov test, which we test the homogeneity of the various alternatives (Table 3). Alternatives B and D does not satisfy the condition of homoscedasticity.

Table.1 Alternatives for One-way ANOVA - Monte Carlo simulation

Alternatives	Normal distribution			Normality	Homogeneity of variance
A	$N_1(10,25)$	$N_2(10,25)$	$N_3(10,25)$	yes	yes
B	$N_1(10, 4)$	$N_2(10,25)$	$N_3(10,49)$	yes	no
C	$N_1(\mu,25)$ $\quad$ $\gamma_3 = 1.15$	$N_2(\mu,25)$ $\gamma_4 = 2$	$N_3(\mu,25)$	no	yes
D	$N_1(\mu,4)$ $\quad$ $\gamma_3 = 1.15$	$N_2(1\mu,25)$ $\gamma_4 = 2$	$N_3(\mu,49)$	no	no

2.2 *Generating random numbers*

Data should be generated for the Monte Carlo simulation. To create simulation models can also use MS Excel and its enhancements: Risk Solver, @Risk, Risk Analyzer, Monte Carlo. In Microsoft Excel for generating random numbers, you can use the command *RAND ()*, we get a random number with uniform distribution in the interval (0,1), or you can use the *"Random Number Generation"* the Data Analysis ToolPak on the Tools menu. We get a random number $X \sim N(\mu, \sigma^2)$ with a given means and standard deviation. The program STATISTICA for generating random numbers, you can

use the *"Rnd (x)"*, which generates a random number of interval (0, x), or the *"RndNormal (x)* ', which calculates.

The number from a normal distribution with a means *0* and standard deviation *x*. Example of generating random number is in Figure 2.
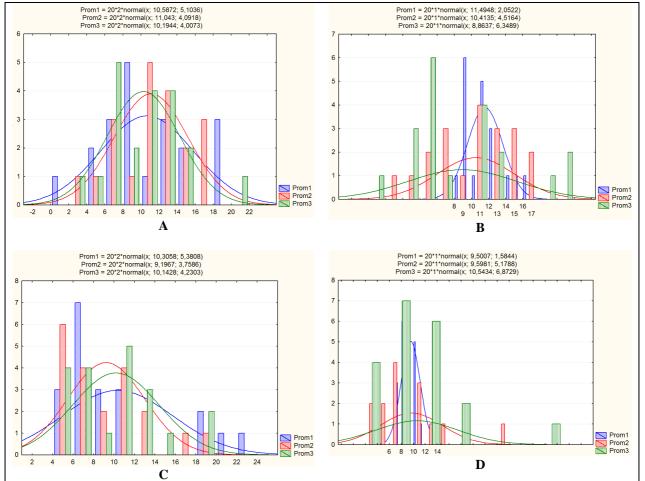

Figure 1 The graph probability density of the alternatives

Table 2 Shapiro-Wilk test

Alternatives	SW-W (A=0.05) 0.90499997	p-value	Result (A=0.05)
A_1	0.973	0.8174	H_0 accepted
A_2	0.9555	0.4587	H_0 accepted
A_3	0.9534	0.4224	H_0 accepted
B_1	0.9309	0.1604	H_0 accepted
B_2	0.9439	0.2834	H_0 accepted
B_3	0.9127	0.0717	H_0 accepted
C_1	0.8074	0.0011	*H_0 rejected*
C_2	0.8964	0.0354	*H_0 rejected*
C_3	0.9238	0.1172	H_0 accepted
D_1	0.9503	0.3722	H_0 accepted
D_2	0.8864	0.0231	*H_0 rejected*
D_3	0.8349	0.0030	*H_0 rejected*

MONTE CARLO SIMULATION FOR ANOVA
Gabriela Ižaríková

Table 3 Levenov test - testing homogeneity of variances

Variable	Levene Test of Homogeneity of Variances (levene) Marked effects are significant at $p < .0500$							
	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
A	31.76869	2	15.88435	481.5002	57	8.447372	1,880389	0,161781
B	*194.5966*	*2*	*97.29828*	*523.8941*	*57*	*9.191124*	*10.58611*	*0.000123*
C	8.288350	2	4.144175	575.9281	57	10.1040	0.410152	0.665492
D	*116.0580*	*2*	*58.02901*	*363.2120*	*57*	*6.372414*	*9.106674*	*0.000370*

Since we need to generate value from $N(\mu, \sigma^2)$ a given means and standard deviation that can be generated directly using the "Random Number Generation" (MS Excel) and enter the parameters or to perform transformations (2):

$$Y = X * \sigma + \mu, \qquad (2)$$

where $X \sim N(0, 1)$ and $Y \sim N(\mu, \sigma^2)$, μ is the means value and is variance.

In the case of data generation with determined skewness and kurtosis it is appropriate to use Fleishman's power of transformation methods. Fleishman's the squares polynomial transformation (3) has the form:

$$Y = a + b * X + c * X^2 + d * X^3, \qquad (3)$$

where Y is the transformed variable with the desired skewness and kurtosis, and X $\sim$ $N(0, 1)$ and *a, b, c, d* are the coefficients of which are, for some pairs of skewness and kurtosis tabulated, for example, we used the values of Table 4.

ANOVA procedure was implemented for the various alternatives and tracks the probability of passing a null hypothesis. The group had of identical means and changed only valid or invalid assumptions about normality and equal variances ANOVA. This means that the null hypothesis should not be rejected. Results simulations (p-value) are in Table 5 and Table 6.
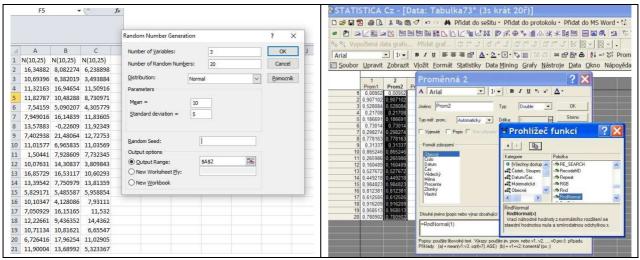

Figure 2 Generating random number

Table 4 Coeficients for Fleishmans transformation

Skewness	Kurtosis	a	b	c	d
1.15	*2*	*-0.1858204*	*0.9368777*	*0.1858204*	*0.0092367*

MONTE CARLO SIMULATION FOR ANOVA
Gabriela Ižaríková

Table 5 Results ANOVA – p-value

Sim.	A		B		C		D	
	p-value		*p-value*		*p-value*		*p-value*	
1	0.9211	H_0 accepted	0.8347	H_0 accepted	0.6609	H_0 accepted	0.7718	H_0 accepted
2	0.8001	H_0 accepted	0.4882	H_0 accepted	0.3202	H_0 accepted	0.8374	H_0 accepted
3	0.2915	H_0 accepted	0.9821	H_0 accepted	0.9458	H_0 accepted	0.6719	H_0 accepted
4	0.5469	H_0 accepted	0.7435	H_0 accepted	0.0237	*H_0 rejected*	0.8086	H_0 accepted
5	0.6912	H_0 accepted	0.2623	H_0 accepted	0.8217	H_0 accepted	0.9492	H_0 accepted
6	0.2072	H_0 accepted	0.7946	H_0 accepted	0.8590	H_0 accepted	0.8237	H_0 accepted
7	0.0128	*H_0 rejected*	0.9091	H_0 accepted	0.8286	H_0 accepted	0.8537	H_0 accepted
8	0.3029	H_0 accepted	0.7089	H_0 accepted	0.0213	*H_0 rejected*	0.2053	H_0 accepted
9	0.7165	H_0 accepted	0.8639	H_0 accepted	0.1920	H_0 accepted	0.1952	H_0 accepted
10	0.2688	H_0 accepted	0.9614	H_0 accepted	0.8598	H_0 accepted	0.7428	H_0 accepted

Table 6 Outup of simulation method Monte Carlo for n=10

	Normality	Equal Var	H_0 accepted	H_0 rejected	Freq.
A	*yes*	*yes*	*9*	*1*	*10%*
B	*yes*	*no*	*10*	*0*	*0%*
C	*no*	*yes*	*8*	*2*	*20%*
D	*no*	*no*	*10*	*0*	*0%*

For each alternative. was found the percentage of rejection of the null hypothesis at a significance level of 5%, Tab.6. In the case of meeting the assumptions of normality were refusals 10% of cases, even if the conditionality correlation variance. When normality was not met we reject the null hypothesis twice if it was met assumption of conformity variances, it means that we have committed type I error in 20% of cases. The results of the simulation study for the 100 simulations are in Table 7, which indicates that the method is sensitive to ANOVA assumption of equal variances as the normality of the data.

Table 7 Outup of simulation method Monte Carlo for n=100

	Normality	Equal Var	H_0 accepted	H_0 rejected	Freq.
A	*yes*	*yes*	*96*	*4*	*4%*
B	*yes*	*no*	*89*	*11*	*11%*
C	*no*	*yes*	*92*	*8*	*8%*
D	*no*	*no*	*84*	*16*	*16%*

Conclusion

The article is an example of Monte Carlo simulations for using ANOVA. It's proven to have the fulfillment of the assumptions of normality of data and correlation scattering on ANOVA results. In the event of failure of assumptions it can be used to compare mean values more than two core set of non-parametric tests, example Kruskal-Wallis test.

References

[1]   FLEISHMAN, A.I.: *Functions for Simulating Data by Using Fleishman's Transformation*, [Online], Available:
https://support.sas.com/publishing/authors/extras/65378_Appendix_D_Functions_for_Simulating_Data_by_Using_Fleishmans_Transformation.pdf
[10 Apr 2016], 2016
[2]   KOČIŠKO,M.: *Simulácia výrobných systémov*, FVT TUKE, 2016. (Original in Slovak)
[3]   MALINDŽÁK, D. et al.: *Modelovanie a simulácia v logistike* /teória modelovania a simulácie/. Košice: TU-BERG, p. 181, 2009. (Original in Slovak)
[4]   TREBUŇA, P. et al: *Modelovanie v priemyselnom inžinierstve*, TUKE, 2015. (Original in Slovak)
[5]   STRAKA, M.: *Diskrétna a spojitá simulácia v simulačnom jazyku EXTEND*, Košice, TU FBERG, Edičné stredisko/AMS, [Online], Available: http://people.tuke.sk/martin.straka/web/web_download/Simulation_scriptum_2.pdf [10 Apr 2016], 2007. (Original in Slovak)
[6]   BOHÁCS, G., SEMRAU, K. F.: Automatic visual data collection in material flow systems and the application to simulation models, *Logistics Journal*, p. 1-7, 2012.

Review process
Single-blind peer reviewed process by two reviewers.