Tạp chí KHLN số 1/2018 (93 - 103)
©: Viện KHLNVN-VAFS
ISSN: 1859 - 0373
Đăng tải tại: www.vafs.gov.vn

UNIVARTIATE, MULTIVARIATE AND PERMUTATIONAL ANALYSIS OF VARIANCE METHODS FOR FORESTRY DATA BY USING R

Bui Manh Hung
International Cooperation Division (ICD), Vietnam National University of Forestry, Xuan Mai, Chuong My, Hanoi

Key words: Anova, ancova, forestry data, manova, peranova, permanova, R language.

ABSTRACT

Analysis of Variance has been widely used in the analysis of forestry research data. They have contributed to a very common question in forest science: are factors affect results of experiments? Analysis of variance (ANOVA) can help researchers to analyze the effect of one or more factors on experimental results. In the meanwhile, ANCOVA besides examining the effects of factors, they also help to check the effect of covariance as well as the relationship between factors and covariance. To make more objective and accurate conclusions, MANOVA should be applied, because MANOVA is able to analyze the effect of factors on experimental results based on various continuous variables. Permutational univariate and multivariate analysis of variance (PERANOVA and PERMANOVA) are new analysis tools. These tools do not require any assumptions. Because of this, scientists can apply them in various fields of forest science. To support and implement PERANOVA and PERMANOVA, the R language should be implemented. The reason is that powerful statistical analysis softwares like SPSS, Stata or Sas is difficult or impossible to conduct these contents.

Từ khóa: Anova, ancova, manova, ngôn ngữ R, peranova, permanova, số liệu lâm nghiệp

Các phương pháp phân tích phương sai đơn biến, đa biến và đa biến lặp số liệu lâm nghiệp bằng R

Phân tích phương sai đã và đang được sử dụng rộng rãi trong phân tích số liệu nghiên cứu Lâm nghiệp hiện nay. Chúng đã góp phần trả lời câu hỏi rất phổ biến trong khoa học lâm nghiệp: các nhân tố có ảnh hưởng tới kết quả thí nghiệm hay không? Phân tích phương sai (ANOVA) đơn biến có thể giúp các nhà nghiên cứu phân tích được ảnh hưởng của một hoặc nhiều nhân tố đến kết quả thí nghiệm. Trong khi đó với ANCOVA, ngoài việc kiểm tra ảnh hưởng của các nhân tố, chúng còn giúp kiểm tra ảnh hưởng của hiệp biến cũng như mối quan hệ giữa các nhân tố với hiệp biến. Để có những kết luận khách quan và chính xác hơn, MANOVA nên được áp dụng bởi lẽ MANOVA có thể phân tích ảnh hưởng của các nhân tố tới kết quả thí nghiệm dựa trên nhiều biến liên tục khác nhau. Phân tích phương sai đơn và đa biến lặp (PERANOVA và PERMANOVA) những công cụ phân tích phương sai mới ra đời. Những phân tích này không yêu cầu bất kỳ điều kiện nào. Cũng vì lẽ đó, các nhà khoa học có thể ứng dụng chúng trong các lĩnh vực khác nhau của khoa học lâm nghiệp. Để hỗ trợ và thực hiện các phân tích phương sai lặp thì ngôn ngữ R nên được áp dụng bởi lẽ hiện nay các phần mềm phân tích thống kê mạnh như SPSS, Stata hay Sas rất khó hoặc không thể thực hiện được những nội dung này.

I. INTRODUCTION

Univariate and multivariate analysis of variance play a very important role in the field of data analysis in general and analysis of forestry data in particular. The British statistician Fisher provided an experimental chart in which concurrent factors were applied, and he was also the first person to contribute to the development of statistical models for such experiments and it is called analysis of variance (Nguyễn Hải Tuất *et al.*, 2006; Jerrold H. Zar, 2010). ANOVA has proven to be an excellent advantages for testing the influence of factors such as light, fertilizer, planting techniques, particle processing techniques, cutting techniques ect. to results of the experiment. These are also tools that help us find the best level in the factor that influence the results of the experiment. Analysis of variance also allows researchers to differentiate between groups such as differences between forest status, among plantation densities or among different tree species. This is a difficult problem and situation when analyzing data. This group of tools can be analyzed either based on a single variable (single-variable analysis) or on multiple independent variables (multivariate analysis) (Roxy Peck and Jay L. Devore, 2012; Jari Oksanen, 2016). For the above reasons and analysis, the use and application of these tools is indispensable in the analysis of forestry research data.

With such importance and indisputable need, but the use and choice of the right tools is a great challenge for researchers. There are many reasons to explain this. At present, with the development of modern statistics, there are many tools for analyzing variance such as univariate analysis (one or many factors) (ANOVA), analysis of covariance (ACOVA), multivariate analysis of variance (MANOVA), multivariate covariance analysis (MACOVA). Recently, with the development of the R language, many new concepts in the analysis of variance have emerged and have been widely applied because of their distinct advantages. They are permutational ANOVA (PERANOVA), permutational ANCOVA (PERANCOVA) and permutational MANOVA (PERMANOVA) and permutational MACOVA (PERMACOVA). A second reason makes the application of the analysis of variance difficult is the sample size is not large enough and assumptions of ANOVA application. Classical analysis of variance requires two assumptions to be satisfied: the samples must have a normal distribution and the variances must be equal. These two conditions are very big barriers when applying ANOVA. They lead to wrong application or turn away of scientific researchers. And one last reason for the difficulty of applying the ANOVA tools is the lack of comprehensive documentation and instruction for the use of all these tools.

In recent years, with the appearance and rapid development of the R language in statistical analysis, many new statistical tools have been developed and implemented easily and quickly. R has been demonstrating and remarkable strengths in the analysis of forestry data. R is an open language and environment for statistical and graphical analysis. R provides and can perform a variety of statistical analyzes such as linear and nonlinear models, tests, time series analysis, clusster analysis, mixed effec models, spatial distribution analysis, etc. (R-Project, 2016). Another advantage of R is the open language. So, there are many packages that can be installed and used in the R environment (Francis Smart, 2014).

For the reasons and the need mentioned above, this paper will (1) analyze systematically how to apply both classical and modern ANOVA tools for analyzing forestry research data with the assistance of the R language, and (2) analyzing the strengths and weaknesses of each tool and (3) propose the recommendations to foresters to select the appropriate analytical tools for their research, especially the multivariate analysis of variance.

II. MATERIAL AND METHODOLOGY

2.1. Material

The used data for analysis in this paper is the diameter and height data collected from five plots in Kon Ka Kinh National Park, Gia Lai Province, Vietnam's Central Highlands. The park location is shown in the following figure (Bui Manh Hung, 2016).

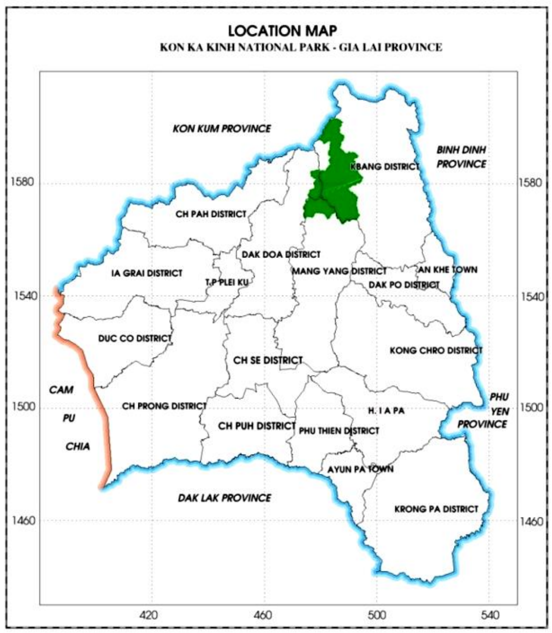

Figure 01. Kon Ka Kinh National Park location

Particularly, the analysis of the two-factor analysis of variance was based on experiments in the nursery. The author examined the effect of 3 shading methods: 0%, 50% and 100% and 3 fertilizer formulas on the growth of *Amoora gigantea* in the nursery. The experiment was arranged in a randomized block (Bui Manh Hung, 2016).

2.2. Selective document and data method

A number of R introduction and multivariate statistical analysis using R documents were collected and analyzed selectively. Materials and analytical works in the forestry sector are a priority. However, the amount of these materials is very limited. The materials were collected and analyzed on the basis of theoretical analysis using R, achievements and results obtained in the field of forestry research data analysis by R.

2.3. Testing and comparison method

From the statistical analysis of commands used to analyze univariate, multivariate and permutational multivariate with the support of packages in R, commands for the forestry data analysis were built elaborately and precisely. Then, these commands were run with the forest datasets. After that, outputs are checked, evaluated and compared with outputs of other softwares such as SPSS, Stata and Sas. Based from those comparison, correct and efficient commands were selected for analyzing univariate, multivariate and permutational multivariate analysis of variance for forestry data.

III. RESULT AND DISCUSSION

3.1. Analysis of variance (ANOVA)

Analysis of variance has been applied in forestry data analysis since the 1960s (Nguyễn Hải Tuất, 1981; Bijan Payandeh, 1970). And then, it has been widely used in various aspects of data analysis such as forest inventory, silviculture, forest soil, forestry biotechnology. Analysis of variance is used to test effects of one or two factors on experimental results. For example, the effect of shading or fertilizer on the diameter and height growth of seedlings in the nursery (Bui Manh Hung, 2011). Or, when analyzing the effect of chemical concentration and substrate on the root length of cuttings. ANOVA is also used to compare several independent samples, such as the comparison of tree diameter in different plots. ANOVA is categorized according to the number of testing factors and covariates.

3.1.1. ANOVA-single factor

Suppose that effects of factor A on the results of the experiment is being tested. When analyzing variance, the following steps are often applied.

- Hypothesis:

H_o: Factor A does not influence data significantly.

H_1: Factor A influences data significantly.

- Calculate sum of square, degrees of freedom, mean square, F test, then complete a analysis of variance table as follows (Bui Manh Hung, 2016; Jerrold H. Zar, 1999):

Table 01. One-way ANOVA results

Source of variation	Sum of square (SS)	Degrees of freedom (DF)	Mean square (MS)	F	F_{05}
Between samples	$SSA=\sum_{i=1}^{k}\sum_{j=1}^{n_i}X_{ij}^2-C$	N-1	$MSA=\frac{SSA}{N-1}$	$F=\frac{MSA}{MSE}$	$F_{0.05(k-1,N-1)}$
Error	$SSE=\sum_{i=1}^{k}\frac{\left(\sum_{j=1}^{n_i}X_{ij}\right)^2}{n_i}-C$	k-1	$MSE=\frac{SSE}{k-1}$		
Total	SSA + SSE	N-k			

-Conclusion

If $F \le F_{0.05(k-1,\ N-k)}$, accept H_o. That means that there is no difference in variables among plots. If $F > F_{0.05(k-1,\ N-k)}$, reject H_o. That means that there is significant difference in variables among plots.

In this article, the effect of a topographic factor on tree diameter growth in 5 different plots was examined. The following commands were used in R.

```
Factor = as.factor(Dia_hinh)
Model = aov(D ~ Factor)
summary(Model)
```

To check normal distribution assumption for plots (for example: Plot 1):

```
OTC1 = subset(Anova, Anova$Dia_hinh
= 1)
shapiro.test(OTC1$D)
```

Analysis results of the topography effect on the diameter growth of natural forest trees by ANOVA are presented in the following table.

Table 02. One way ANOVA results

```
           Df Sum Sq Mean Sq F value Pr(>F)
Factor       4    627  156.67   1.657  0.158
Residuals  976  92272   94.54
```

The above results show that the effect of terrain is not significan, because the Pr value is greater than 0.05. In other words, the diamter growth of forest trees in plots is similar.

3.1.2. ANOVA: two-factor

Suppose that the effect of two factors A and B on the results of the experiment is being tested, so two-factor ANOVA will be used. Two-factor analysis of variance includes two cases: with and without replications. Replication means that in each level of factor A and each level of factor B has more than or equal to 2 values. Opposite, without replication means that each combination has only one value.

The steps of analyzing the two-way ANOVA are quite similar to single-factor ANOVA. However, two-way ANOVA (with replications) can check an interactionn between two factors A and B. Two-way ANOVA table is summerized as follows (Roxy Peck and Jay L. Devore, 2012).

Table 03. Two-factor ANOVA

Source	SS	DF	MS	F test	F_{05}
-Factor A - Factor B - Error - Total	V_A V_B V_N V_T	a-1 b-1 (a-1)(b-1) n-1 = (ab-1)	$S_i^2(A)$ $S_i^2(B)$ $S^{"2}$	$\frac{S_i^2(A)}{S^{"2}}$ $\frac{S_i^2(B)}{S^{"2}}$	k_1 = a-1 k_2 = (a-1)(b-1) k_1 = b-1 k_2 = (a-1)(b-1)

Conclusion: If $F \leq F_{05}$ then factors A and B do not affect the results of the experiment. In contrast, factors A and B have a significant effect on the data of the experiment.

Analysis of variance can be conducted to test the effect of 1, 2 or more factors on the results of the experiment. Foresters can also analyze the interaction between factors, the effect of each pairs and influence all factors. However, in this paper, analysis of variance is limited to 2 factors of light and fertilizer to seedling diameter. The number of replications are 2. Following commands are used in R.

```
as = as.factor(Anh_sang)
pb = as.factor(Phan_bon)
Model = aov(D ~ as + pb + as*pb)
summary(Model)
```

The analysis results of the two factors effect (light and fertilizer) on seedling diameter in nursery are shown in the next table.

Table 04. Two-way ANOVA results

```
           Df Sum Sq Mean Sq F value   Pr(>F)
as          2 118.20   59.10  57.381 7.54e-06 ***
pb          2   2.03    1.01   0.984    0.411
as:pb       4  13.53    3.38   3.284    0.064 .
Residuals   9   9.27    1.03
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From above results, only light factor has a significant effect on the growth of seedlings in the nursery. In contrast, the fertilizer factor does not have a significant influence (Pr value greater than 0.05). Besides, light and fertilizer do not interact with each other.

3.2. Analysis of Covariance (ANCOVA)

Analysis of covariance is an extension of the analysis of variance. ANCOVA not only analyzes effects of factors on the results of experiments, but also allows the foresters to control the effects of covariates on the experimental results. Analysis of covariance uses both linear regression analysis and ANOVA. Dependent variables are continuous variables, independent variables can be both continuous and discrete variables (Nguyễn Văn Tuấn, 2006). This is a difference from the one-way ANOVA. When doing ANCOVA, scientists can know the effect of factors and covariates on the results of experiments, and can find out a relation between factors and covariates. These are also outstanding advantages of analysis of covariance.

If y is a tested variable such as diameter, height. A is an affecting factor, for instance: forest type, light and fertilizer. Here, assume that the influencing factor has two groups 1 and 2. A covariate is x, for example: soil properties, pH ect. Therefore, a basic linear regression model of ANCOVA for groups 1 and 2 would be (Nguyễn Văn Tuấn, 2006):

$$y_1 = \alpha_1 + \beta.x + e_1$$

$$y_2 = \alpha_2 + \beta.x + e_2$$

Which:

α_1 is average of y when x = 0 for group 1.

α_2 is average of y when x = 0 for group 2.

β is the slope of the relation between x and y.

e_1 and e_2 are errors with mean 0 and variance σ^2.

The regression model of covariance can be combined as follows:

$$y = \alpha + \beta.x + \gamma.g + \delta.(xg) + e$$

The model can be explained as follows: The dependent variable y (diameter, height) depends on the covariate x (soil property), the influencing factor A (forest type) and the interaction between x and factor A.

In a similar way to the analysis of variance, ANCOVA can be performed to test the effect of more than two factors, as well as an interaction between them on results of experiments. In this paper, ANCOVA single factor is used to test the effect of topography on tree diameter, along with effects of a covariate (pH of the soil) in different terrain locations. The following commands are called in R.

```
Factor = as.factor(Dia_hinh)
Model = aov(D~Dia_hinh+pH+Dia_hinh*pH)
summary(Model)
```

In forestry data for example, in addition to examining the effect of topography on tree diameter, one more factor to check is the pH of the soil at plots. The comparison will give better results, if the pH is the same between the plots. Thus, the pH becomes covariate in this case. The result of covariance analysis is as follows:

Table 05. Analysis of Covariance results

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Dia_hinh	1	0	0.1	0.001	0.9750	
pH	1	600	600.2	6.353	0.0119	*
Dia_hinh:pH	1	0	0.1	0.001	0.9811	
Residuals	977	92298	94.5			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The results show that topography does not affect diameter growth significantly. However, the pH has a significant influence on diameter and between the pH and the terrain, there is no interaction.

In the analysis of research data in general and forestry data in particular, disadvantages of ANOVA as well as ANCOVA are that they are possible to examine effects of factors and covariates on only one dependent variable (eg diameter, height, humidity, root length, etc.). To solve this problem, in mathematical statistics, there is a commonly used tool which is multivariate analysis of variance.

3.3. Multivariate analysis of variance (MANOVA)

Multivariate analysis of variance is simply the analysis of variance, but for many dependent variables. Thus, ANOVA will examine the difference between means of two or more groups, when MANOVA examines the difference between mean vectors of groups (Jari Oksanen, 2016). And for that reason, multivariate variance analysis can be applied to one factor, two factors or more.

Within the limits of this paper, mathematical theories and calculations of SS, DF and MS for MANOVA will not be presented, to avoid repeatations. In contrast, the applicability of multivariate analysis is further analyzed. MANOVA should be applied more than ANOVA if experiments of the study have two or more continuous dependent variables, because the multivariate advantage of MANOVA will lead to a more accurate and objective conclusions about the effect of factors on the study object.

In forestry research, measured and tested individuals are often trees, probably seedlings, or timber trees in forests. Therefore, when examining effects of factors such as fertilizer or light on seedling growth in the nursery. Or checking the effect of planting density, slope, age or forest types on tree growth, instead of just checking the effect of those factors on only diameter or only height, it is much better to test effects on diameter, height and event biomass of the trees at the same time. As a result, conclusions of forestry scientists will be more comprehensive, solid and accurate about the influence of external factors on the growth of seedlings in the nursery or on the trees living in the forest.

In this paper, the effect of topography on both diameter and height is analyzed. From that, there are more precise conclusions about the effect of topography on forest tree growth. The following commands are run in R.

```
Factor = as.factor(Dia_hinh)
Y <-cbind(D, H)
Model <-manova(Y ~ Factor)
summary(Model)
```

Obtained results are as follows.

Table 06. Multivariate analysis of variance results

```
           Df   Pillai approx F num Df den Df    Pr(>F)    
Factor      4 0.038583   4.7997      8   1952 7.266e-06 ***
Residuals 976                                              
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The above results indicate that if examining two variables diameter and height simultaneously, the topography has a significant effect on forest tree growth, because the Pr value is less than 0.05. This conclusion is contrary to the conclusion drawn from ANOVA for only diameter variable. The main reason is that the height variable is also considered and included in the model. Thus, it can be said that MANOVA can evaluate the impact of topography on the tree growth more comprehensively and objectively.

However, a weakness of all the above analysis tools is assumptions. All univariate and multivariate ANOVA are parametric methods. Therefore, they require many conditions to be satisfied such as: populations must be normal distributions, population variances must be equal and samples must be independent. Considering only the following assumption: populations must be normal distributions. Now checking used dataset to know is that satisfied that assumption? For example, Shapiro-Wilk test results in R for plot 1 are as follows.

Table 07. Results of normal distribution checking for plot 1

```
Shapiro-Wilk normality test

data:  OTC1$D
W = 0.75873, p-value = 3.519e-16
```

The results of the above table illustrate that a diameter frequency distribution of trees in plot 1 is not close to the normal distribution, even it is very different from the normal. This is evidenced by the very small p-value.

As a result, it can be said that the statistical analysis above is not usable. Results of the impact of topography on forest plant growht are not reliable and accurate to conclude. In other words, the results of univarate and multivariate analysis of variance should not be used. In the past, for situations like these, forestry scientists have encountered many difficulties in order to continue to analyze.

3.4. Permutational univariate and multivariate analysis of variance (PERANOVA and PERMANOVA)

In recent years, with supports of the R language, permutational methods and tools have been developed and applied more and more widely. However, in the field of forestry research in Vietnam, these tools are quite strange, especially PERANOVA and PERMANOVA. However, the exploitation and application of these tools in the analysis of forestry research data should be developed and disseminated. The reason is that permutational analysis of variance will overcome the assumption disadvantage of classical ANOVA tools.

Both univariate and multivariate ANOVA and ANCOVA require data to be satisfied by two main assumptions: populations must be normal distributions and population variances must be equal. Regarding to the first assumption, it is a great obstacle to properly apply the analysis of variance. The reason is that many forest research data do not be close to the normal distribution. For example, tree diameter frequency distributions of natural forests, especially secondary forests. They usually have a decreasing distribution or a J-shape distribution. They are not asymptotic to the normal distribution (Bui Manh Hung, 2016). Another example is diameter/height or species frequency distributions regenerating trees in natural forests. These distributions are almost always a declining distribution. So if foresters apply the univariate and multivariate ANOVA and ANCOVA in these cases, this is completely incorrect. Results of the analysis are not accepted statistically.

In addition, there are many other cases where distributions are not really close to the normal distribution such as diameter/height frequency distributions of young even age plantation tres and so on. In those cases, the normal distribution tests such as Q-Q plot, Kolmogorov-Smirnov test, Shapiro-Wilk test will reject a null hypothesis. So, the question is what should foresters going to do in these situations? In the past, scientists often used a method of transforming data by taking logarithms or taking square roots. However, these methods have a disadvantage that the data will be changed, so that analysis results are no longer the result of the original data. The result belongs to the converted data. As a consequence, analysis results do not really make much sense. In previous years, forestry researchers were really stuck in these situations.

Today, a new solution has been found to overcome this difficulty, which is the permutational analysis of variance. The dominant advantage of permutational analysis of variance and permutational multivariate analysis of variance is that there are no assumptions for the data. Because they are non-parametric statistical tools (Andreas Hamann, 2016). Therefore, whether our data has a normal distribution, it is still possible to perform the analysis of variance and the analysis results are meaningful. These analyses can be carried out relatively easily in the R language (Jari Oksanen, 2016).

PERMANOVA uses distance matrices for partitioning sum of square. To calculate pseudo-F test statistic, the following formula is applied (Bui Manh Hung, 2016; Marti J. Anderson, 2011).

$$F = \frac{SS_A/(a-1)}{SS_W/(N-a)}$$

Where: $SS_A = SS_T - SS_W$

In which:

SS_T = sum of squared distances divided by N (total number of observations).

SS = sum of squared distances within groups divided by n (total number of observations per group).

a = the number of groups.

This is more visual in the following figure.

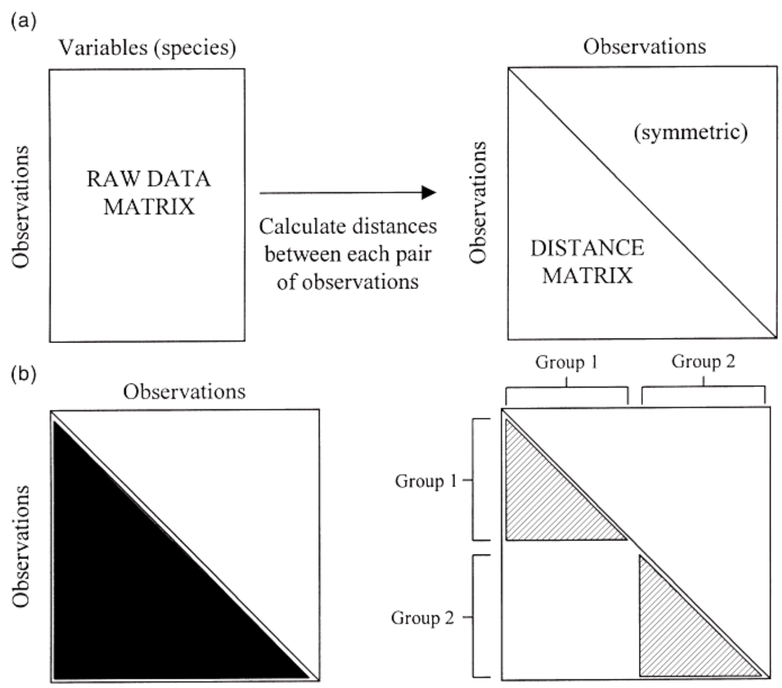

Figure 02. A diagram for the computation of (a) a distance matrix from a raw data matrix and (b) a non-parametric MANOVA statistic (Marti J. Anderson, 2011)

In this article, an effect of topography on tree diameter growth was again examined, but PERANOVA was applied. This overcomes ANOVA's limitations. Used R command are:

```
Factor = as.factor(Dia_hinh)
Model = aovp(D ~ Factor)
summary(Model)
```

For PERMANOVA analysis, the study analyzes the effect of the topographic factor on both diameter and height variables of forest trees. The used command is as follows.

```
adonis(Permanova ~ Dia_hinh, Factor)
```

The following is PERANOVA results. They show that the terrain does not significantly affect growth (Pr value is greater than 0.05). In other words, the tree diameter is homogeneous among plots. A great thing here is that the result is contrary to the ANOVA result. The reason is that the diameter frequency distribution is not normal.

Table 08. Permutational analysis of variance results

```
            Df R Sum Sq R Mean Sq Iter Pr(Prob)
Factor       4      627   156.667 2415  0.06791 .
Residuals  976    92272    94.541
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

To improve an objectivity of testing the effects of topography on plant growth. PERMANOVA was performed for both diameter and height variables. And the interesting result is that topography affects forest tree growth significantly (Pr = 0.109). This shows an opposite conclusion to above PERANOVA. That is the precious value of multivariate analysis, which gives scientists more accurate conclusions, closer to the natural world. These analyses are currently only possible in R, which can not be performed by other powerful statistical softwares such as SPSS, Stata or Sas.

Table 09. Permutational analysis of variance results

```
Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

           Df SumsOfSqs  MeanSqs F.Model     R2 Pr(>F)
Dia_hinh    1     0.111 0.110991  2.6473 0.0027  0.109
Residuals 979    41.047 0.041927         0.9973
Total     980    41.158                  1.0000
```

IV. CONCLUSION

The analysis of variance tools have been widely used in the analysis of forestry research data, because these tools have proved undeniable strengths, especially they are very suitable for a frequent research question in a forestry sector: are factors affect experiment results?

In this article, advantages and application of each ANOVA tool have been presented and analyzed in detail. Univariate ANOVA can help researchers to analyze the effects of one or more factors on the experiment results. In other words, the data between groups is uniform or not homogenous (Nguyễn Hải Tuất *et al.*, 2006). The above example, the effect of topographic factors on tree diameter growth is not significant. Regarding to ANCOVA, in addition to examining the effects of factors, they also examine the effect of covariance as well as the relationship between factors and covariates (Nguyễn Văn Tuấn, 2006). In the example in the result part, the effect of pH is very significant on the diameter growth, and there is no relationship between pH and topography. For more objective and accurate conclusions, MANOVA should be applied, because MANOVA can analyze the effect of factors on experimental results based on various continuous variables (Jari Oksanen, 2016). Therefore, for cases which have many independent variables, forestry scientists should use MANOVA to make findings more persuasive and objective. However, all ANOVA tools (both univariate and multivariate) are limited by assumptions: populations must be normal distributions, population variances must be equal and samples must be independent. If one of the applicaton conditions is not met, then the analysis results are no longer valid, and researchers will fall into a deadlock.

To solve this problem, PERANOVA and PERMANOVA should be applied. They are new tools for analyzing variances and are gradually being applied because of their advantages. These analyses do not require any assumptions, because they are non-parametric analyses (Jari Oksanen, 2016; Andreas Hamann, 2016). They are based on repeated data extraction. For this reason, scientists can safely and reliably apply them in various fields of forest science such as forest inventory, forest ecology, silviculture, forest soil, forest resources management and forest product processing. Therefore, they should be applied and diffused.

To support and implement permutational analyses, the R language should be used. Nowadays, powerful statistical analysis software like SPSS, Stata or Sas are difficult or impossible to do that. In addition, the R language is open sources with many packages, so it is useful to help foresters in order to find solutions for many real deadlock situations with classical statistical analysis tools.

REFERENCES

1. Nguyễn Hải Tuất, Vũ Tiến Hinh, Ngô Kim Khôi, 2006. Phân tích thống kê trong lâm nghiệp. Nhà xuất bản Nông nghiệp, Hà Nội, Việt Nam.
2. Jerrold H. Zar, 2010. Biostatistical Analysis (5th Edition*)*: Prentice Hall, Upper Saddle River, New Jersey 07458, USA.
3. Roxy Peck and Jay L. Devore, 2012. Statistics: the exploration and analysis of data: Brooks/Cole Cengage Learning, 20 Channel Center Street, Boston, MA 02210, USA.
4. Nguyễn Văn Tuấn, 2006. Phân tích số liệu và tạo biểu đồ bằng R.: NXB Khoa học và Kỹ thuật, Hà Nội, Việt Nam.
5. Jari Oksanen, 2016. Package 'vegan'. Department of Biology, University of Oulu, P.O.Box 3000, 90014 Oulu, Finland. Available from: https://cran.r-project.org/web/packages/vegan/vegan.pdf (Accessed 27 April, 2016).
6. R-Project, 2016. What is R?. Available from: https://www.r-project.org/about.html (Accessed 09 February, 2016).
7. Francis Smart, 2014. Why use R? Five reasons. Michigan State University, Available from: http://www.econometricsbysimulation.com/2014/03/why-use-r-five-reasons.html (Accessed 09 February, 2016).
8. Bui Manh Hung, 2016. Structure and restoration of natural secondary forests in the Central Highlands, Vietnam, in Chair of Silviculture, Institute of Silviculture and Forest protection, Faculty of Environmental Sciences. Dresden University of Technology.
9. Nguyễn Hải Tuất, 1981. Thống kê toán học. Nhà xuất bản Nông nghiệp, Hà Nội, Việt Nam.
10. Bijan Payandeh, 1970. Comparison of Methods for Assessing Spatial Distribution of Trees Forest Science, 1970(16): p. 312 - 317.
11. Bui Manh Hung, 2011. The influence of light, planting methods and densities on the growth of Amoora gigantea and Pygeum arboreum in the North of Vietnam, in Fenner School. The Australian National University.
12. Jerrold H. Zar, 1999. Biostatistical Analysis. Prentice Hall, Upper Saddle River, New Jersey 07458, USA.
13. Andreas Hamann, 2016. Permutational ANOVA and permutational MANOVA. Department of Renewable Resources, Faculty of Agricultural, Life, and Environmental Sciences, University of Alberta, Canada. Available from: https://www.ualberta.ca/~ahamann/teaching/renr480/Lab13.pdf (Accessed 27 April, 2016).
14. Marti J. Anderson, 2011. A new method for non-parametric multivariate analysis of variance. Austral Ecology, 2011(26): p. 32 - 46.

Email của tác giả chính: hungbm@vfu.edu.vn
Ngày nhận bài: 12/12/2017
Ngày phản biện đánh giá và sửa chữa: 03/03/2018
Ngày duyệt đăng: 05/03/2018