STATISTICAL ANALYSIS APPLICATION IN CLASSICAL AND POPULAR MUSIC

Tomas Ruzgas,

Department of Applied Mathematics
Kaunas University of Technology
Lithuania.

ABSTRACT

Independence between discrete variables and categorical factors is analysed, where discrete variables are notes, note types and rests. Identity of factor style which can obtain values classic, romantic or modern can be recognized by analysis of variance or chi-square test. Grouping of these variables by clustering methods for each mode (major and minor) is presented. As a result some clusters match musical triads or chords. Interpretation of clusters is shown by dendrogram with dividing line. The results of note type clusters demonstrate the effect of modern style, where the syncopated rhythm dominates.

In this paper we investigate the problem how to choose the number of clusters. In addition we try to answer the question if the probability clustering is more advantageous than the geometrical clustering. The results show that mathematical statistics methods produce output that satisfies the norms and standards of music.

Keywords: Music Data Analysis, Analysis of Variance, Homogeneity Groups, Probability Clustering, Geometrical Clustering, Number of Clusters.

INTRODUCTION:

What is music? What is the sound of music? These questions are probably never going to find a definitive answer. The exact definition is outside the scope of this work. For example, modern scientists are interested to establish the specific acoustic characteristics allowing people to distinguish one musical instrument from the other (Hartman, 1997). New questions and some interesting discoveries were obtained while trying to answer this question. Music is the only art area where the shape and means correspond; just the same as mathematics is the only science where methods and subjects are identical. The relationship between mathematics and music is the result, i.e. the product of thinking without a tangible form (Blagoveshchenskaya, 2004). Pythagoras was the first man who found a connection between mathematics and music. The sound coherence laws were discovered and this provided the basis for the modern diatonic music. An instrument tuning named Pythagoras is still used today. Music sounds are in harmony if it satisfies certain mathematical conditions. Despite tight connection between music and mathematics, the idea of using mathematical methods in musicology was used only in the 1950s. Last decade various mathematical methods of mathematical statistics, such as Fourier harmonic analysis, Markov chain models was successfully applied in musiclogy. Though, stochastic methods such as correlation analysis, time series analysis, analysis of variance, data clustering were used very rarely in this area.

Let $X(1), X(2), ..., X(n)$ be a sample of independent random values (musical notes) from compositions, chosen by factor of style, composer or mode. This discrete variable can get a whole number value from 0 to 11, which means that octaves are divided into twelve equal semitones (Blagoveshchenskaya, 2004). So all the piano keys factorized modulo 12 represent a cyclic structure in the sense that the notes sound in unison (we would say, coincide) if and only if the visible distance between their keys is equal to an octave that is just 12 keys. Notes in an equal temperament octave form an abelian group with 12 elements (Lewin, 1982). $\langle Z_{12};+\rangle$ – commutative group is used to prepare the data for analysis by transformations and alterations. Other variables – rests and types of notes – can gain discrete values from $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, ..., \frac{1}{64}\}$ and $\{1, 1/2^*, 1/4^*, 1/8^*, ..., 1/64^*\}$, where $1^* = 1 + 1 \cdot 0.5$, $1/2^* = \frac{1}{2} + \frac{1}{2} \cdot 0.5, ...$. It is important that analysis is made for transposed compositions to C-major and a-minor and values are independent of their order in sample.

This paper comprises the following sections: Section 2 describes methods and models used for music works analysis; Section 3 reviews the analysed estimators and summaries results; Section 4 contains the resumptive conclusions.

METHODS AND MODELS:

The data consist of 330 compositions in classic, romantic and modern style. Information about variables and factors is kept in music **.xml* files. After information of musical compositions is analysed and musical norms and standards are evaluated, several methods of statistical analysis are suggested:

a) reports and graphs of descriptive statistics;
b) goodness of fit hypothesis test of *notes*, *note types* and *rests* by factors;
c) analysis of variance:
   - factors – musical *style* and *composer* – influence on means of *notes*, *note types* and *rests*,
   - Tukey's multiple grouping for *styles* and *composers*;
d) correlation analysis and contingency tables statistics:
   - level of relationship between *notes*, *note types* and *rests*;
   - relationship between categorical factors and variables;
e) clustering:
   - *notes* and *rests* classification by *mode* and *style*, when geometrical and probability methods are employed.

Review of main these methods and results are presented.

Analysis of variance displays whether there is relation between dependent variables and factors. In analysis of variance Tukey's multiple grouping method considers all possible pairwise differences of means at the same time. Suppose there are n observed *notes* (or *note types*, or *rests*) $X(1), ..., X(n)$ from an approximate normal distribution with mean μ and variance σ^2 and studentized range is g. The Tukey confidence limits for all pairwise comparisons with confidence coefficient of at least $(1–\alpha)$ are (Tukey, 1953; Kramer, 1956):

$$\bar{X}_i - \bar{X}_j \pm \frac{1}{2} g_{\alpha;r;n-r} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}, \; i,j=1,\dots,r \;\; i \neq j \qquad (1)$$

with number of *style* (or *composer*) groups r, total number of data values n and size of groups n_i and n_j (Rafter, Abell, & Braselton, 2002).
Independence of categorical variables can also be identified by chi-square test. The chi-square statistic is computed as (Radlow & Alf, 1975)

$$Q = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}}, \qquad (2)$$

where $e_{ij} = \dfrac{n_i n_j}{n}$, n_{ij} – expected frequency of i^{th} value of *notes* (or *note types*, or *rests*) in j^{th} category.
When grouping variable values by there frequencies the results are clusters. We use three types of clustering methods: geometrical clustering and probability ones. Geometrical clustering joins clusters with the smallest distances, which can be evaluated by (Sokal & Michener, 1958; Milligan, 1980; Ward, 1963):

- average linkage method – $D_{KL} = \dfrac{1}{n_K n_L} \sum_{i=1}^{n_K} \sum_{j=1}^{n_L} d(X(i), X(j))$; in average linkage the distance between two clusters is the average distance between pairs of *notes* (or *rests*), one in each cluster;
- centroid method – $D_{KL} = \left\| \bar{X}_K - \bar{X}_L \right\|^2$; in the centroid method, the distance between two clusters is defined as the (squared) Euclidean distance between their centroids or means;
- single linkage method – $D_{KL} = \min_{i \in C_K} \min_{j \in C_L} d(X(i), X(j))$; in single linkage, the distance between two clusters is the minimum distance between the *notes* (or *rests*) in one cluster and the *notes* (or *rests*) in the other cluster;
- Ward's minimum-variance method – $D_{KL} = \dfrac{\left\| \bar{X}_K - \bar{X}_K \right\|^2}{1/n_K + 1/n_L}$; in Ward's minimum-variance method, the distance between two clusters is the ANOVA sum of squares between the two clusters added up over all the variables; at each generation, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous generation; and others.

The special case of geometrical clustering is procedure that combines an effective method for finding initial clusters with a standard iterative algorithm for minimizing the sum of squared distances from the cluster means. The result is an efficient procedure for disjoint clustering of large data sets. This procedure was directly inspired by the *k-means algorithm* (MacQueen, 1967). It uses a method that Anderberg M. R. calls *nearest centroid sorting* (Anderberg, 1973). A set of points called *cluster seeds* is selected as a first guess of the means of the clusters. Each observed *note* (or *rest*) is assigned to the nearest seed to form temporary clusters. The seeds are then replaced by the means of the temporary clusters, and the process is repeated until no further changes occur in the clusters.
Notes (or *rests*) clustering using the EM algorithm. If the density function of the random vector X has q maxima, it can be approximated by a mixture of q unimodal densities:

$$f(x) = \sum_{k=1}^{q} p_k f_k(x). \qquad (3)$$

Let the distribution of the random vector X depend on a random variable v that takes on the values 1,...,q. It is interpreted as the number of class the observed object belongs to. In the classification theory quantities $p_k = \mathbf{P}\{v=k\}$ are called *a priori* probabilities that the observed object belongs to the k^{th} class, and quantities $\pi_k(x) = \mathbf{P}\{v=k|X=x\}$ are *a posteriori* probabilities. The function f_k is treated as the conditional density of X as $v=k$. By the term soft clustering of a sample we refer to the estimation of the values $\pi_k(X(t))$ for all $k=1,...,q$, $t=1,...,n$. A sample is hard-clustered if estimators $\hat{v}(1),..., \hat{v}(n)$ of $v(1),\ldots, v(n)$ are indicated where $v(t)$ denotes the class number of the *note* (or *rest*) $X(t)$.
The mixture of Gaussian distributions is the most popular model in the clusterisation theory and practice. Therefore, in this section, we assume that $f_k(x)$ are Gaussian densities with means M_k and covariance matrices R_k. Let $f(\theta,x)$ denote the right-hand side of equation (3), where $\theta = ((p_k, M_k, R_k)\ k=1,\ldots,q)$. Since

$$\pi_k(x) = \frac{p_k f_k(x)}{f(\theta, x)}, \quad k = \overline{1,q}. \qquad (4)$$

the estimators of *a posteriori* probabilities are obtained as usual by the "plug-in" method which replaces the unknown parameter vector θ on the right side of (4) by its maximum likelihood estimate $\theta^* = \arg\max_{\theta} L(\theta)$, $L(\theta) = \prod_{t=1}^{n} f(\theta, X(t))$. The EM algorithm, an iterative procedure most frequently used to find this estimate, was also applied in this study.

Let $\hat{\pi}_k = \hat{\pi}_k^{(r)}$ be the estimates obtained after r cycles of the iterative procedure. Then a new estimate $\hat{\theta} = \hat{\theta}^{(r+1)}$ is defined by the equalities

$$\hat{p}_k = \frac{1}{n}\sum_{t=1}^{n} \hat{\pi}(X(t)),$$

$$\hat{M}_k = \frac{1}{n\hat{p}_k}\sum_{t=1}^{n} \hat{\pi}(X(t)) \cdot X(t),$$

$$\hat{R}_k = \frac{1}{n\hat{p}_k}\sum_{t=1}^{n} \hat{\pi}(X(t)) \cdot [X(t) - \hat{M}_k] \cdot [X(t) - \hat{M}_k]'$$

for all k=1,…,q. Inserting $\hat{\theta}^{(r+1)}$ into the right-hand side of expression (4) we get $\hat{\pi}^{(r+1)}(X(t))$, $k = \overline{1,q}$, $t = \overline{1,n}$. Using this iterative procedure a non-decreasing sequence $L(\hat{\theta}^{(r)})$ it obtained, however its convergence to the global maximum depends on the initial value $\hat{\theta}^{(0)}$ (or $\hat{\pi}^{(0)}$). The simplest solution of the initial value selection problem is a random start technique: the EM procedure is repeated many times from the random starting value $\hat{\pi}^{(0)}$. The result with the maximal value of $L(\hat{\theta})$ is selected as final. The methodology of consecutive extraction of the mixture components (Rudzkis & Radavicius, 1995) can be also applied as well.

For choosing the number of clusters q, various tests of model adequacy can be used. Beginning with q = 1, the values of the parameter q are consecutively increased untill the hypothesis on model (3) adequacy is not rejected (e.g., with the significance level α = 0.1). To test this hypothesis, a criterion based on the increment of the maximum likelihood function can also be applied. Let $\hat{\theta}(q)$ be an estimator of θ obtained by the EM algorithm with the number of clusters equal to q. Let

$$\psi = L(\hat{\theta}(q+1)) - L(\hat{\theta}(q)) \quad (5)$$

and let $G(u)$ denote an estimated distribution of ψ obtained by making use of the parametric bootstrap (see, for example, Hall, 1992; Wong, 1985) under the condition that mixture model (3) of Gaussian distributions is valid. Then the hypothesis on model (3) adequacy is not rejected if

$$1 - G(\psi) \geq \alpha. \quad (6)$$

The number of clusters obtained in this way is denoted by q^*.

Applying hierarchical techniques it is possible to obtain a dendrogram (binary tree) whose terminal nodes are the treatment means (Rienzo, Guzman & Casanoves, 2002).

SUMMARY OF RESULTS:

Analysis of variance and correlation shows that styles differ by independent variable *note*. Analysis of variance as a result gives homogeneity groups by Tukey grouping method (see Table 1), which excludes styles into separate groups, where N – is size of sample, with it's mean. Coefficient of determination indicates that the model accounts for less than 1% of variation. But in case data consist of 330 compositions, it's a small part of population in all different styles.

Table 1. Grouping of variable *note* by factor *style*.

Tukey Grouping	Mean	N	Style
A	0.179210	114108	Romantic
B	0.163817	140008	Classic
C	0.148669	249005	Modern

Correlation analysis statistics show that variable *note* and factor *style* has no relation with probability 0.0001, which means that they are dependent (see Table 2). Rests and type of notes gives similar results in analysis of independence.

Table 2. Chi-square test for variable *note* and factor *style*.

Statistic	DF	Value	Prob
Chi-Square	22	2744.99	<.0001
Likelihood Ratio Chi-Square	22	2728.32	<.0001
Mantel-Haenszel Chi-Square	1	29.9297	<.0001

Nonparametric *k*-means clustering results for compositions in major mode shows better results than other, because the smallest number of members in cluster is chosen to be 2. The biggest frequency has third cluster, which has tonic, dominant and the leading down degree in music scale (see Table 3).

Table 3. Clusters of notes in major.

CLUSTER	Notes			
1	D+	G+	A+	.
2	C+	E	F	A
3	C	D	G	.
4	F+	B	.	.

In minor mode *k*-means clustering gives minor tonic triad with subdominant degree of scale in the first cluster, which is most frequent cluster in compositions with minor mode (see Table 4). If first cluster is the most frequent, then composer has more minor compositions than major. A personal analysis can be made by analysing Figure 1. Bizet, Debussy and Gershwin have more minor compositions in this sample.

Table 4. Clusters of notes in minor.

CLUSTER	Notes			
1	C	D+	F	G
2	C+	F+	A	.
3	D	G+	.	.
4	E	A+	B	.

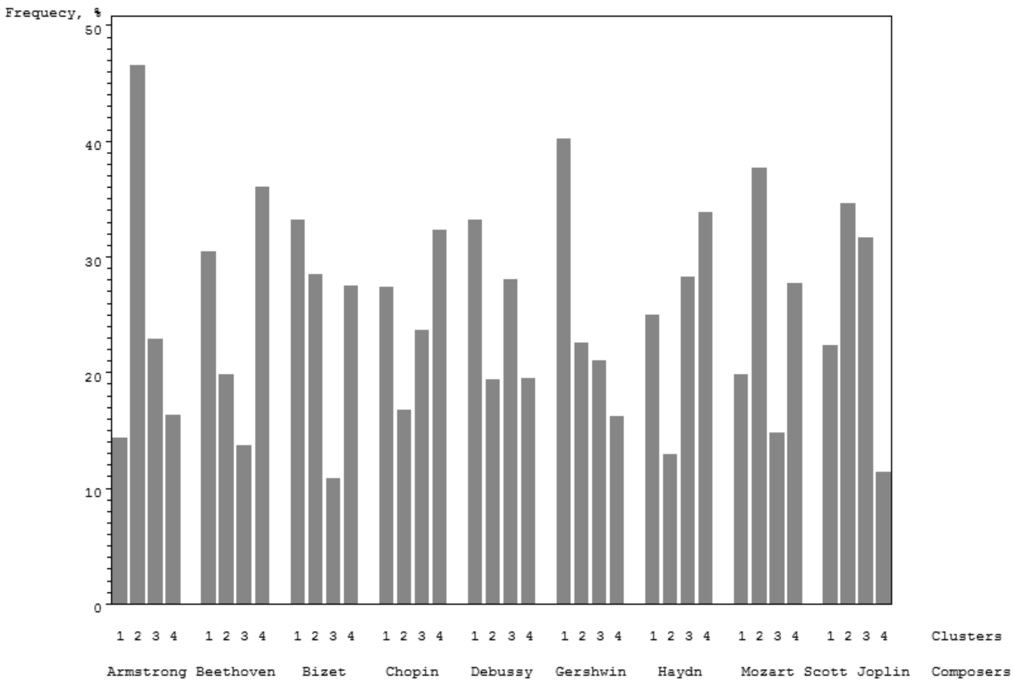

Figure 1. *Note* clusters' from table 4 frequencies.

In the dendrogram, we have clusters of variable – *note type*, which gives similar results as variable rest in minor mode compositions. There are used geometrical single linkage, average linkage, centroid and Ward's methods, and probability the EM algorithm for clustering. Geometrical methods can be easily realized and they are widely used of its simple application (Lampropoulos & Tsihrintzis, 2004; Novello, McKinney & Kohlrausch, 2011). The average linkage tends to join clusters with small variances, and it is slightly biased toward producing clusters with the same variance. A disadvantage of Ward's algorithm is that the method does not guarantee optimal partitioning of objects into clusters. Moreover, due to the natwe of clustering, the minimum value is contingent on previously formed clusters, somewhat biasing the results. The centroid method is more robust to outliers than most other hierarchical methods but in other respects might not perform as well as Ward's method or average linkage. The single linkage method shows good cluster separation in the studied data, therefore the largest distinctness deserves the EM algorithm. Additionally, this algorithm has more accurate results. Vertical line is the dividing line, where criterion ψ consider choosing the clusters. The dependency on the size of music works, the type of clustering, the number of clusters, and the distance between the investigated objects is of rather a similar nature (in term of quality). The result is two multiple clusters: first one has general values 1/2, 1/4, 1/16, second – 1/64, 1/4*, 1/2*, 1/32, 1/16*, 1/8*, 1 (see Figure 2). Values that make a single-value cluster are 1/32*, 1*, 1/8. Clustering with different distance method, we get different clusters results if we choose the same dividing line.

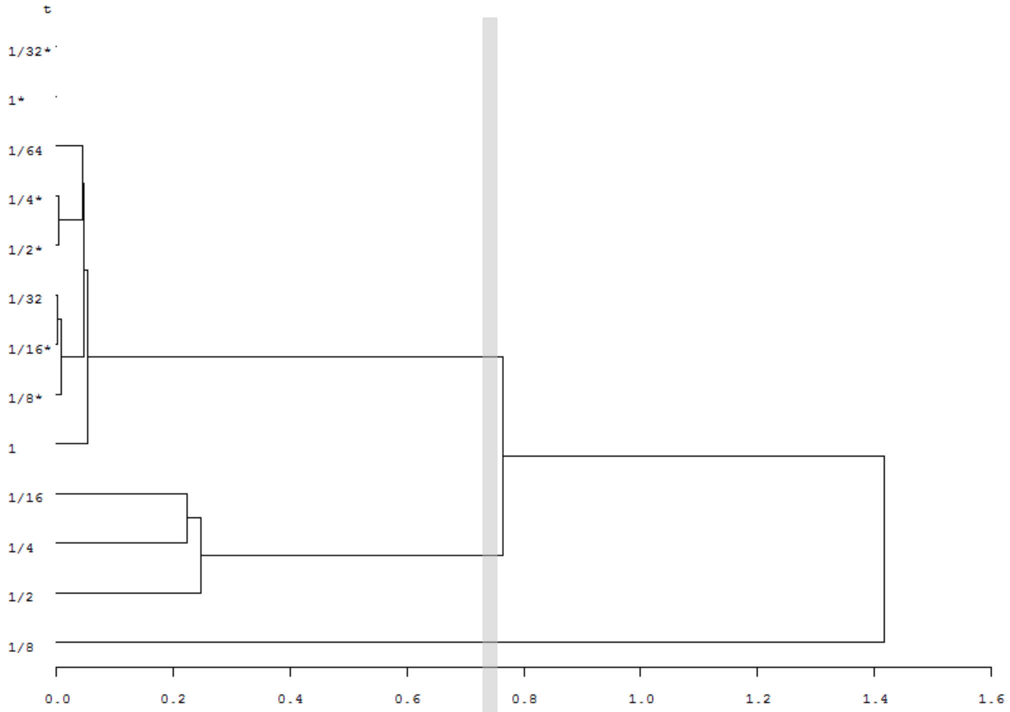

Figure 2. Dendrogram of *note type*.

Difference in rhythm between current styles is shown in Table 5. All the shorter notes, that are prolonged, in modern style appear with bigger probability. It illustrates syncopated rhythm, which is more popular in jazz and other modern genres.

Table 5. Distribution of variable *note type* by *style*.

t	Classic, %	Romantic, %	Modern, %
1/64	0	0.02	0
1/32	0.84	0.99	8.48
1/32*	**0.27**	**0**	**2.18**
1/16	22.91	18.35	29.17
1/16*	**1.01**	**1.53**	**7.88**
1/8	50.15	45.8	31.43
1/8*	**1.64**	**1.8**	**4.13**
1/4	15.23	21.89	9.01
1/4*	**1.31**	**2.06**	**1.11**
1/2	5.19	5.87	3.94
1/2*	**0.39**	**1.38**	**0.31**
1	1.05	0.32	2.34
1*	**0**	**0.01**	**0.02**

CONCLUSIONS:

The investigation made by classical, romantic or modern music styles and their composers has proved the point that style or composers identity can be recognized by statistical analysis results. The clustering analysis of notes creates chords or triads that satisfy the norms and standards of music – analysis of minor compositions excludes the tonic triad. Cluster analysis visually shows differences between styles and composers. Rests and types of notes are clustered into groups of classical and syncopated rhythm, so the information about composition style is characterized.

REFERENCES:

[1] Anderberg, M. R. (1973). Cluster Analysis for Applications, *New York: Academic Press*.
[2] Blagoveshchenskaya, E. (2004). Some links between Music and Mathematics – algebraic aspects. *Essener Kollegs für Geschlechterforschung, 4(2)*, 4-22.
[3] Hall, P. (1992). The bootstrap and edgeworth expansion, *New York: Springer*.
[4] Hartman, W. M. (1977). Signals, Sound, and Sensation. *AIP Press*, *Woodburg*, *New York*.
[5] Kramer, C. Y. (1956). Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications. *Biometrics, 12*, 307–310.
[6] Lampropoulos, A. S., Tsihrintzis, G. A. (2004). Agglomerative Hierarchical Clustering For Musical Database Visualization and Browsing. *Proc. 3rd Hellenic Conference on Artificial Intelligence*, 5-8.
[7] Lewin, D. (1982). A Formal Theory of Generalized Tonal Functions. *Journal of Music Theory, 26(1)*, 23-60.
[8] MacQueen, J. B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 281–297.
[9] Milligan, G. W. (1980). An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms. *Psychometrika, 45*, 325–342.
[10] Novello, A., McKinney, M. M. F., Kohlrausch, A. (2011). Perceptual Evaluation of Inter-song Similarity in Western Popular Music. *Journal of New Music Research*, *40(1)*, 1-26
[11] Radlow, R., Alf, E. F. (1975). An Alternate Multinomial Assessment of the Accuracy of the Chi-Square Test of Goodness of Fit. *Journal of the American Statistical Association, 70*, 811–813.
[12] Rafter, J. A., Abell, M. L., Braselton, J. P. (2002). Multiple comparison methods for Means. *Society for Industrial and Applied Mathematics, 44(2)*, 259-278.
[13] Rameau, J. P. (1725). A Treatise of Music, Containing the Principles of Composition London, translated from the French *Traité de l'harmonie réduite à ses principes naturels*, *Ballard*, *Paris*, *1722*.
[14] Riemann, H. (1895). Harmony Simplified, *Augener, London, translated from the German Vereinfachte Harmonielehre by Rev. H. Bewerunge*.
[15] Rienzo, J. A., Guzman, W. A., Casanoves, F. (2002). A Multiple-Comparisons Method Based on the Distribution of the Root Node Distance of a Binary Tree. *American Statistical Association and the International Biometric Society Journal of Agricultural, Biological, and Environmental Statistics*, *7(2)*, 129-142.
[16] Rudzkis, R., Radavicius, M. (1995). Statistical Estimations of a Mixture of Gaussian Distributions. *Acta Applicandae Mathematicae, 38,* 37-54.
[17] Schenker, H. (1930). Das Meisterwerk in der Musik, *Drei MaksenVerlag*, *Munich*, *3 volumes*, *published 1925, 1926, 1930*.
[18] Sokal, R. R., Michener, C. D. (1958). A Statistical Method for Evaluating Systematic Relationships. *University of Kansas Science Bulletin, 38*, 1409–1438.
[19] Tukey, J. W. (1953). The Problem of Multiple Comparisons. *The Collected Works of John W. Tukey*, *Vol 8, 1994*, *New York: Chapman & Hall*.
[20] Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*, 236–244.
[21] Wong, M. (1985). A bootstrap testing procedure for investigating the number of subpopulations, *Journal of Statistical Computation and Simulation*, *22*, 99-112.
