The finitely geometric symptom analysis in the glioma survival study

Nina Alexeyeva
Dep. of Math.and Mech.
State University
St.Petersburg, Russia
ninaalexeyeva@mail.ru

Ivan Smirnov
Dep. of Math.and Mech.
State University
St.Petersburg, Russia
ivan.b.smirnov@gmail.com

Polina Gracheva
Lab. of Math. Modeling
Federal Almazov HBE centre
St.Petersburg, Russia
PolinaGracheva@yandex.ru

Boris Martynov
Dep. of Neurosurgery
Military Medical Academy
St.Petersburg, Russia
omartynova2005@rambler.ru

Abstract

Novel approach for statistical inference in survival analysis based on factor or dichotomic variables is proposed. We are seeking for the most informative finitely linear combinations (symptoms) of variables in the finite field. This procedure necessarily yields the new variable of the same nature (e.g. factor). Different measures can be used as an optimality criterion of such combination: entropy, the uncertainty coefficient, p-level of some statistical tests. We use this method to determine the major factors of the glioma postoperation survival. These factors were found to be age, high-grading factor, stage of illness and a factor of the stereotactic cryodestruction.

1. Introduction

The usual way to deal with factor variables with two levels is to treat them as dichotomic and study their linear combinations in the real field. Following [1], [2] we propose different approach. Let the number of factor levels equal to a prime number power. Then the factor variable itself can be viewed as the element of some Galois field, thus it becomes possible to study finitely linear combinations of factors in such field. This idea can be applied to dichotomic variables and field F_2 is appropriate here. Then finitely linear combination of dichotomic variables yields a new variable of the same kind with some special meaning.

For example, let variable X_1 be equal to 1 if height of object of interest is big and 0 otherwise; variable X_2 be equal 1 if weight is huge and 0 otherwise. Then, dichotomic variable $X_{12} = X_1 + X_2(mod\ 2)$ represents the inadequacy between height and weight. Variable X_{12} derived this way can be more informative for statistical analysis.

2. The symptom analysis

2.1. The symptom and syndrome definitions

First, note that the values of the dichotomic vector $X = (X_1, \ldots, X_n)^T$ are the elements of the finite Euclid geometry $EG(n, 2)$ and components $X_i,\ i = 1, 2, \ldots, n$ are elements of the dual projective geometry $PG(n-1; 2)$ [3].

Definition 1. *For $\tau = (t_1, \ldots, t_k) \subseteq (1, 2, \ldots, n)$ and $A_\tau = (a_1, \ldots, a_n)^T$ with components $a_j = 1$ if $j \in \tau$ and $a_j = 0$ otherwise define **symptom** X_τ to be a finitely linear combination of the form*

$$X_\tau = A_\tau^T X(mod\ 2). \tag{1}$$

Hence symptoms X_τ can be viewed as the points of the projective geometry $PG(n-1; 2)$. Every symptom is given its own number in binary code as follows:

$$X^m = X_\tau,\ m = a_1 2^0 + a_2 2^1 + \ldots + a_n 2^{n-1}. \tag{2}$$

The number of non-nil coefficients a_i means *the symptom rank*. Symptoms with a single non-nil coefficient a_i are *trivial* $X^{2^{k-1}} = X_k$, other are *nontrivial*. The *singular* symptom $X_\emptyset$ equals zero with probability one.

Remark. *$X^m + X^m = X_\emptyset(mod\ 2)$ and $X^m + X_\emptyset = X^m(mod\ 2)$ for all m.*

Definition 2. *The set of all $2^n - 1$ symptoms X^m from* (2) *forms **syndrome** Δ_{n-1} of order $(n-1)$.*

A finitely combination of any pair of symptoms in the syndrome belongs to the same syndrome. A single symptom X_τ can be viewed as syndrome Δ_0. Two symptoms X_τ and X_μ belong to the syndrome $\Delta_1 = (X_\tau, X_\mu, X_\nu)$, where $\nu = \tau \triangle^1 \mu$. The syndrome Δ_k can be obtained via induction in accordance with the *counting sequence* as

$$\Delta_k = (\Delta_{k-1}, X_\nu, \Delta_{k-1} + X_\nu(mod\ 2)), \tag{3}$$

where $X_\nu \notin \Delta_{k-1}$. Symptoms in the counting sequence with $2^k - 1$ raiting are *basic*. Symptoms that don't form by pairs the syndrome with order 1 can be used as basic.

2.2. Syndromes and block designs

Definition 3. ***Block design** is an incidence system $D(v, b, r, k, \lambda)$ in which a set X of v points is partitioned*

1. The triangle sign denotes the symmetric set difference operation throughout the paper

α	$X_\alpha = 0$	$X_\alpha = 1$
(1)	0,2,4,6	1,3,5,7
(2)	0,1,4,5	2,3,6,7
(12)	0,3,4,7	1,2,5,6
(3)	0,1,2,3	4,5,6,7
(13)	0,2,5,7	1,3,4,6
(23)	0,1,6,7	2,3,4,5
(123)	0,3,5,6	1,2,4,7

Table 1. The block design $D(8, 14, 7, 4, 3)$ with elements from $\Omega_3 = \{0, 1, \ldots, 7\}$ in the binary code.

into a family of b subsets (blocks) in such a way that any two points determine λ blocks with k points in each block, and each point is contained in r different blocks.

In accordance with the Zinger's theorem [3] the projective geometry $PG(n; 2)$ corresponds to the block design $D(2^{n+1} - 1, 2^{n+1} - 1, 2^n - 1, 2^n - 1, 2^{n-1} - 1)$.

The block design $D(v, b, r, k, \lambda)$ with $v = b$ and $k = r$ is symmetric and will be denoted as $D(v, k, \lambda)$. This, if we take symptoms of the syndrome Δ_n as elements and syndrome Δ_{n-1} as blocks, we will obtain symmetric block design $D(2^{n+1} - 1, 2^n - 1, 2^{n-1} - 1)$.

The automorphism group $SL_{n+1}^{F_2}$ of this block design is the automorphism group of the syndrome Δ_n too.

Denote by Ω_n set of all possible values of random vector $X = (X_1, \ldots, X_n)^T$, $X_i \in \{0; 1\}$. Any symptom X^m induce the partition $\Omega_n = B_m \cup \bar{B}_m$, where $B_m \cap \bar{B}_m = \emptyset$ and $|B_m| = |\bar{B}_m| = 2^{n-1}$; here m is as in (2). Points of the basic block B_m and additional block $\bar{B}_m$ satisfy equations $X_\tau = 0$ and $X_\tau = 1$ accordingly. The singular symptom $X_\emptyset$ induces the partition $(\Omega_n, \emptyset)$. $v = 2^n$ points of Ω_n and $b = 2(2^n - 1)$ blocks $(B_\tau, \bar{B}_\tau)$ form a block design $D(v, b, r, k, \lambda)$ corresponding to the finite Euclid geometry $EG(n, 2)$ where $r = 2^n - 1,\ k = 2^{n-1},\ \lambda = 2^{n-1} - 1$.

For example, for $n = 3$ the syndrome Δ_2 contains symptoms $X_\tau, X_\mu, X_{\tau\mu}, X_\nu, X_{\tau\nu}, X_{\mu\nu}, X_{\tau\mu\nu}$ which are elements of $D(7, 3, 1)$. Syndrome Δ_1 is constructed from triples of symptoms: $(X_\tau, X_\mu, X_{\tau\mu})$, $(X_\tau, X_\nu, X_{\tau\nu})$, $(X_\tau, X_{\mu\nu}, X_{\tau\mu\nu})$, $(X_\mu, X_\nu, X_{\mu\nu})$, $(X_\mu, X_{\tau\nu}, X_{\tau\mu\nu})$, $(X_{\tau\mu}, X_{\mu\nu}, X_{\tau\nu})$ and $(X_{\tau\mu}, X_\nu, X_{\tau\mu\nu})$ which are blocks of design $D(7, 3, 1)$. It is easy to construct the partitions of Ω_3 which correspond to these symptoms and form $D(8, 14, 7, 4, 3)$ (for example in Tab.1 at $\tau = (1)$, $\mu = (2)$, $\nu = (3)$). For the sake of simplicity, elements of Ω_3 are shown in the binary code $X_1 + 2X_2 + 4X_3$.

2.3. Generalized symptoms and the permutation group of $EG(n, 2)$

Definition 4. *Consider the set partition $\Omega_n = B \cup \bar{B}$ with $B \cap \bar{B} = \emptyset$ and $|B| = |\bar{B}| = 2^{n-1}$. The generalized symptom in the syndrome Δ_{n-1} is a new dichotomic variable Z with value zero when $X = (X_1, \ldots, X_n)^T \in B$ and value one otherwise.*

Let Z_1 and Z_2 be generalized symptoms in Δ_{n-1} and

$$|B_1 \cap B_2| = |\bar{B}_1 \cap B_2| = |B_1 \cap \bar{B}_2| = |\bar{B}_1 \cap \bar{B}_2| = 2^{n-2}.$$

Then the generalized symptom Z_{12} equals to 0 when $X \in B_1 \cap B_2 \cup \bar{B}_1 \cap \bar{B}_2$ and equals to 1 when $X \in B_1 \cap \bar{B}_2 \cup B_1 \cap \bar{B}_2$. Generalized symptoms Z_1, Z_2 and Z_{12} forms the generalized syndrome with order 1. The generalized syndrome of greater order can be defined inductively via counting sequence similar to usual syndromes. Taking the permutation group of $EG(n, 2)$ into the consideration we obtain different generalized syndromes. In particular, for $n = 3$ they can be described with 30 nonisomorphic permutations of block design $D(8, 14, 7, 4, 3)$ on the base of group $SL_4^{F_2}$ (which is isomorphic to the even permutation group A_8 [4]).

2.4. Informative and stochastic characteristics of syndromes

The syndrome distribution is defined as joint distribution of basic symptoms. It is not difficult to show the invariance of the syndrome distribution towards its automorphism group. The main characteristic of discrete distribution $X : (p_1, \ldots, p_N)$ is the entropy

$$H_X = -\sum_{i=1}^{N} p_i \log_2 p_i. \tag{4}$$

The uncertainty coefficient can be used for the measurement of influence of random variable X to Y:

$$J_{X|Y} = \frac{H_X + H_Y - H_{X,Y}}{H_Y}.$$

Here $H_{X,Y}$ denotes the entropy of joint distribution of X and Y and conditional entropy $H_{X|Y}$ is defined as

$$H_{X|Y} = H_{X,Y} - H_Y, \tag{5}$$

The value $I(X, Y) = H_X + H_Y - H_{X,Y} = H_X - H_{X|Y} = H_Y - H_{Y|X}$ is the *joint information* for X and Y. The next theorem is proved in [2] and can be applied to generalized symptoms as well.

Theorem 1. *For basic symptoms $X_0, X_1, \ldots, X_n \in \Delta_n$ we have $H_{\Delta_n} = H_{X_0} + \sum_{j=1}^{n} H_{Z_j|\Delta_{j-1}}$, where $Z_j = (X_j + X_{\tau_j}(mod\ 2))$, $X_{\tau_j} \in \Delta_{j-1} \cup X_\emptyset,\ j = 1, \ldots, n$.*

Definition 5. *Let X be a factor variable with levels $A = \{x_0, \ldots, x_k\}$ and $\{N_0, \ldots, N_k\}$ is the corresponding frequency set, $\sum_{i=0}^{k} N_i = N$ is the sample size.* ***Variety*** $V_X(A)$

of the random sample on the basis of the variable X with level set A is

$$V_X(A) = N\log_2 N - \sum_{i=0}^{k} N_i \log_2 N_i.$$

It is not difficult to show the connection between the entropy and the variety:

$$V_X(A) = N\widehat{H}_X, \tag{6}$$

where $\widehat{H}_X$ is calculated according to (4) with the empirical distribution $(\frac{N_0}{N}, \dots, \frac{N_k}{N})$.

The variety on the basis of $\mathbf{X} = (X_1, \dots, X_n)^T$ is defined as $I_{X_1}(A) + \dots + I_{X_n}(A)$ [5].

Let $\mathbf{X} = (X_1, \dots, X_n)^T$ be a random dichotomic vector and $\Omega_n = \{0, 1, 2, \dots, 2^n - 1\}$ is the its set of all possible values. Denote by $N_0, \dots, N_{2^n-1}$ the frequency set in the binary code. Consider $C = (c_0, \dots, c_l) \subseteq \Omega_n$ and a dichotomic variable Y which equals to 0 if $\mathbf{X} \in C$ and 1 otherwise. Then we have

$$V_Y(\Omega_n) = g(\Omega_n) - g(C) - g(\bar{C}), \tag{7}$$

where $g(C) = \left(\sum_{i=0}^{l} N_{c_i}\right) \log_2 \left(\sum_{i=0}^{l} N_{c_i}\right)$.

Theorem 2. *Let the generalized symptom Y induce the set partition $\Omega_n = C \cup \bar{C}$ and $|C| = |\bar{C}|$. Then*

$$V_Y(C) + I_Y(\bar{C}) = N \sum_{i=1}^{n} \widehat{H}_{X_i|Y}. \tag{8}$$

Proof: Denote by $(B_i, \bar{B}_i)$ the set partition of $\Omega_n = B_i \cup \bar{B}_i$ induced by the basic symptom $X_i,\ i = 1, \dots, n$. Then from (7) we deduce $V_{X_i}(C) = g(C) - g(CB_i) - g(C\bar{B}_i), V_{X_i}(\bar{C}) = g(\bar{C}) - g(\bar{C}B_i) - g(\bar{C}\bar{B}_i)$.

$$\begin{aligned} V_Y(C) + I_Y(\bar{C}) = \sum_{i=1}^{n} (g(C) - g(CB_i) - g(C\bar{B}_i) + \\ + g(\bar{C}) - g(\bar{C}B_i) - g(\bar{C}\bar{B}_i) + g(\Omega_n) - g(\Omega_n)) = \\ \overset{(7)}{=} \sum_{i=1}^{n} V_{YX_i}(\Omega_n) - V_Y(\Omega_n) = \\ \overset{(6)}{=} N \sum_{i=1}^{n} (\widehat{H}_{YX_i} - \widehat{H}_Y) \overset{(5)}{=} N \sum_{i=1}^{n} \widehat{H}_{X_i|Y}. \end{aligned}$$

□

Via summing of all varieties of all pairs of blocks in design we obtain a variety of block design. Note that lower block design variety means the greater information significance of the generalized symptom Y towards the basic symptoms.

Definition 6. *Let Δ_n consist of $N = 2^{n+1} - 1$ symptoms $X^1, \dots, X^N$. The syndrome **entropy** $H(\Delta_n)$ is the entropy of the joint distribution of the basic symptoms. The summary syndrome entropy $H_\Sigma(\Delta_n)$ is $\sum_{i=1}^{N} H_{X^i}$.*

Theorem 3. *Denote by $c = \frac{1}{2} \sum_{i=1}^{7} H(\Delta_1(i))$, $\Delta_1(i) \in \Delta_2$.*

$$Then \sum_{\tau \subseteq \{1,2,3\}} H_{X_\tau} = \frac{1}{6} \sum_{\tau,\mu \subseteq \{1,2,3\}} I(X_\tau X_\mu) + c. \tag{9}$$

Proof: For symptoms $X_\tau, X_\mu, X_{\tau\mu}$ from the syndrome Δ_1 we have $I(X_\tau, X_\mu) = H_{X_\tau} + H_{X_\mu} - H(\Delta_1)$, $I(X_\tau, X_{\tau\mu}) = H_{X_\tau} + H_{X_{\tau\mu}} - H(\Delta_1)$, $I(X_{\tau\mu}, X_\mu) = H_{X_{\tau\mu}} + H_{X_\mu} - H(\Delta_1)$. Thus $I(X_\tau, X_\mu) + I(X_\tau, X_{\tau\mu}) + I(X_\mu, X_{\tau\mu}) = 2(H_{X_\tau} + H_{X_\mu} + H_{X_{\tau\mu}}) - 3H(\Delta_1)$. Using the characteristics $v = 7$ and $r = 3$ of block design $D(7, 3, 1)$ we obtain

$$3\sum_{i=1}^{v} H(\Delta_1(i)) = 2r \sum_{\tau \subseteq \{1,2,3\}} H_{X_\tau} - \sum_{\tau,\mu \subseteq \{1,2,3\}} I(X_\tau, X_\mu),$$

and (9) follows trivially. Let us show that c is constant. The Δ_1 distribution is determined by probabilities on four components of the vector

$$(B_\tau, \overline{B}_\tau) \otimes (B_\mu, \overline{B}_\mu) = (B_\tau B_\mu, B_\tau \overline{B}_\mu, \overline{B}_\tau B_\mu, \overline{B}_\tau \overline{B}_\mu).$$

Syndromes of $\Delta_1(i)$ exhaust all $28 = C_8^2 = 4 \cdot 7$ possible pairs of elements in $D(8, 14, 7, 4, 3)$ which is formed from symptoms $X_\tau,\ \tau \subset \{1, 2, 3\}$. Hence $\sum_{i=1}^{7} H(\Delta_1(i))$ is same for all permutations of this block design. □

This statement is proved in special case $n = 2$, but can be easily generalized. In that way the minimal summary syndrome entropy $H_\Sigma(\Delta_n)$ means the minimal connection between symptoms.

2.5. Algorithm for determination of the most informative syndromes

Our main aim is to choose the most informative syndromes from the syndrome with higher order. For example, in order to pick out the most informative Δ_2 we should enumerate all possible triples of symptoms forming Δ_2 from the counting ordered Δ_n. The straightforward way is to review all possible $C^3_{2^{n-1}}$ combinations from Δ_2. But the computation complexity of such procedure is inaccessible even for small number of variables (e.g. $n \geqslant 7$).

The main idea of computational complexity reduction is to choose all possible three symptoms forming different syndromes but at the same time exclude permutations of each other since the syndrome entropy is permutation-invariant.

Theorem 4. *Let i and j be the numbers in Δ_n of symptoms X_τ and X_μ forming $\Delta_1 = (X_\tau, X_\mu, X_{\tau\triangle\mu})$. Without loss of generality we might take $i < j$. Then there exists such k that $i < j < k$ and $X^k = X_{\tau\triangle\mu}$ if for some odd $q \geqslant 0$*

$$j \in [b_i + qs_i; b_i + (q+1)s_i), \tag{10}$$

for $b_i = 2^{\lfloor \log_2(i) \rfloor + 1}$ and $s_i = 2^{\lfloor \log_2(i) \rfloor}$

Proof: Denote by X^i a symptom on i-th place in Δ_n. Each pair (i,j) forms new $\Delta_1 = (X^i, X^j, X^k)$ if $i < j < k$ and

$$X^j \neq X^i + X^{j'}(mod\ 2)$$

for every $j' < j$. For the sake of simplicity we will omit the $(mod\ 2)$ sign later on. Also the inclusion $X^i \in \Delta_m$ will always denote that $X^i \in \Delta_m$, $X^i \notin \Delta_{m-1}$ and $X^i \notin \Delta_{m+1}$. Our proof will consist of two steps.

1) Suppose that $X^i \in \Delta_m$. Then pair (i,j) cannot form the syndrome Δ_1 yet unseen during the enumeration if $j < b_i = 2^{m+1}$. Indeed, using the principle of inductive construction of syndromes we obtain

$$X^i = X^{2^m} + X^{i'}, \quad X^j = X^{2^m} + X^{j'}$$
$$\Rightarrow\ X^k = X^i + X^j = X^{i'} + X^{j'},$$

here $X^{i'} \in \Delta_{m-1}$ and $X^{j'} \in \Delta_{m-1}$. Then $X^k \notin \Delta_m$ and $k < j$. This means that symptom constructed on (i,j) is a permutation of $\Delta_1 = (X^k, X^i, X^j)$.

2) Suppose that $i \in [2^m; 2^{m+1})$ and

$$j \in [2^{m+1} + q2^m; 2^{m+1} + (q+1)2^m),$$

where q is odd and

$$1 \leq q < \frac{2^n - 2^{m+1}}{2^m} = 2^{n-m} - 2,$$

Let us show that every pair (i,j) constructed this way won't lead to Δ_1 not seen before.

The same inductive argument can be used to show that if in this case $X^j \in \Delta_{m+1}$ then

$$2^{m+1} + 2^m \leq j < 2^{m+1} + 2^{m+1},$$

thus $X^j = X^{j'} + X^{2^{m+1}}$ for some $j1$m such that $2^m \leq j' < 2^{m+1}$. But $X^i + X^{j'} = X^{k'}$, where $0 \leq k' < 2^m$ and we have $X^k = X^{k'} + X^{2^{m+1}}$, hence

$$2^{m+1} \leq k < 2^{m+1} + 2^m$$

and $i < k < j$ and the syndrome based on (i,j) is a permutation of $\Delta_1 = (X^i, X^j, X^k)$.

Suppose that for every j such that for $k > 1$ $X^j \in \Delta_{m+k}$ all Δ_1 based on (i,j) were already seen. Assume that $X^j \in \Delta_{m+k+1}$. One can easily see that then j belongs to

$$[2^{m+k+1} + 2^{m+1} + q2^m; 2^{m+k+1} + 2^{m+1} + (q+1)2^m) =$$
$$= [2^{m+1} + 2^m(2^{k+1} + q); 2^{m+1} + 2^m(2^{k+1} + q + 1)) =$$
$$= [2^{m+1} + q'2^m; 2^{m+1} + (q'+1)2^m),$$

where q' is odd. Since $i \in [2^m; 2^{m+1})\ then 2^{m+1} = 2^{\lfloor \log_2(i) \rfloor + 1} = b_i$ and $2^m = 2^{\lfloor \log_2(i) \rfloor} = s_i$ the only possibility for j is $j \in [b_i + (q-1)s_i; b_i + qs_i)$. □

Now we are ready to present the algorithm which enumerates all triples of symptoms excluding the permutations of each other.

Step 1. In accordance with (10) for each i we construct the set $W(i)$ of possible variants for j.

Step 2. From (3) we have $\Delta_2 = (\Delta_1, X_\nu, \Delta_1 + X_\nu(mod\ 2))$. Let k denote the number of X_ν. We apply the formula (10) to j substituted for i and k substituted for j in order to obtain set $W'(j)$ of all possible variants for k.

3. The medical application of the symptoms and syndromes in suvival analysis

3.1. Dataset

We analysed the survival data of 280 patients with the mean age of 44 ± 1. 79 patients of them had the protoplasmic astrocytoma (PA), 104 patients had the anaplastic astrocytoma (AA), 97 patients had the glioblastoma multiforme (GBM). The surgical tumor removal, the stereotactic biopsy and the stereotactic cryodestruction were performed for 197, 16 and 67 patients correspondingly. A set of explanatory variables was available as well: a consciousness level, cephalalgias, presence of convulsions, tumor sizes, types of the previous treatment, etc.

3.2. Methods

The Pearson chi-square test and the uncertainty coefficients were used to analyze contingency tables. The Gehan's Wilcoxon (GW) test was used for comparison of multiple suvival curves. Different groups were selected with the help of symptom analysis. They were interpreted by means of stepwise discriminant analysis. Kruskal-Wallis test, t-test of homogenity of variables, Mann-Whitney U test indicated significant difference between the groups obtained (all p-levels were less than 0.05).

3.3. Results

Special computer program was used to extract the most informative syndrome Δ_2 for the maximal difference between suvival curves and a set of 3 most informative symptoms were obtained. Two of them were trivial and means a cephalalgia (presented or not) and the speech disorder as well as one symptom which means an incompatibility between a consciousness level (distinct or no) and convulsions. This way we obtained eight groups of patients denoted later on by $0, 1, \ldots, 7$ which were separated the most by suvival curves. Next we pick out the most informative for survival generalized symptom $(0456, 1237)$ which equals to 0 when a patient belongs to groups $0, 4, 5, 6$ and equals to 1 otherwise. Then we pick out the most informative for survival generalized syndrome $\Delta_1 = (04, 56, 17, 23)$ which is formed by symptoms $(0456, 1237)$ and $(0234, 1567)$ (Fig. 1). Third informative symptom $(0126, 3457)$ is obtained as most informative for survival from eight symptoms which are acceptable for Δ_2.

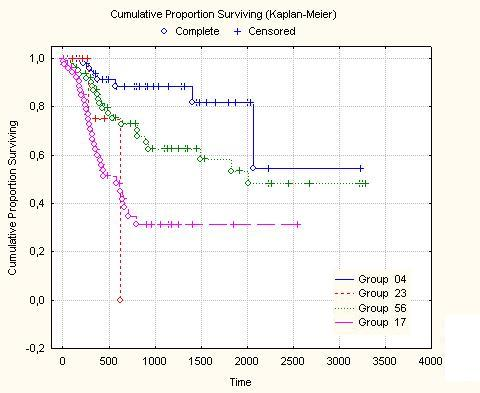

Figure 1. Survival curves for different groups.

Analysing $\Delta_1 = (04, 56, 17, 23)$ and using the stepwise discriminant analysis we singled out the group 23 where 14 patients were older (63 ± 2 in comparison with 43 ± 1). These patients were observed for not so long time (< 126 days) but the rest patients had median survival of 2048 days.

The most successful patients of group 04 were young (41 ± 2 in comparison with 46 ± 1), their tumor size was smaller (41280 ± 6143 in comparison with 58595 ± 4405, $p = 0.03$). The lower survival quartile (LSQ) was 1568 days. The "unfortunate" group 17 with LSQ of 279 days showed bigger mean tumor size (64880 ± 6201 in comparison with 47613 ± 4419, $p = 0.02$).

Analysing generalized symptoms we obtained the high-grading symptom $(0456, 1237)$ ($p = 0.00001$ in GW-test) since in 1237-group (LSQ of 282 days in comparison with 899 days in remaining group) there were 14.67% patients with PA, 36.63% with AA and 58.33% with GBM, the Pearson Chi-square test significance is $p < 10^{-4}$ and the uncertainty is 10%.

The generalized symptom $(0234, 1567)$ ($p = 0.00033$ in GW-test) can be considered as a stage factor since patients of 0234-group (LSQ of 1483 days in comparison with 370 days for remaining group) differed by greater Karnofsky performance score (KPS): 73.5 ± 1.5 compared to 67 ± 1.

Using the minimal variety of the block design permutation at the choice of the third generalized symptom we obtained $(0126, 3457)$ ($p = 0.00112$ in GW-test) that can be considered as the factor of the stereotactic cryodestruction (SC): in 0126-group (LSQ equals to 310 days) only 14% patients had SC operation. In 3457-group (LSQ of 810 days) 30% patients had SC (p-level equals to 0.003).

4. Conclusion

The symptom-syndrome method is proposed for the detection of main survival factors and their properties by means of minimal numbers of variables. This method does not require complete data for all variables in comparison with the Cox proportional hazards model and is more automated in comparison with the recursive partitioning analysis [6].

Acknowledgment

The authors would like to thank Solntsev V.N. for the help with data analysis, Korobeynikov A.I. for some comments on the presentation of paper and Alexeyev A.O. for the symptom analysis support in the view of logical principles and forms of logics [7].

References

[1] N. Alexeyeva and A. Alexeyev, *About a role of the finite geometries in the dichotomic variables correlation analysis*. Mathematical models. Theory and applications. Issue 4. Eds. prof. Chirkov M.K. Research Institution of Mathematic and Mechanic, Saint-Petersburg University. 2004. P. 102-117. (in Russian)

[2] N. Alexeyeva, *The combinatorial analysis of two forms latent periodicities in categorial sequences*. "Vestnik of Petersburg University ", Ser.1. Vol. 3. 2006. P.55-64.(in Russian)

[3] M. Hall, *Combinatorial theory*. Blaisdell, 1967.

[4] A. Bart, N. Klochkova (Alexeyeva), V. Nekrutkin, *Symmeties in the factor analysis structures of stomatological data*. Statistical methods in clinical tests. Saint-Petersburg University. 1999. P.165-219. (in Russian)

[5] *Statistical Methods for Digital Computers.* Eds. by K. Enslein, A. Ralston, H.S. Wilf. John Wiley and Sons. Inc. 1977.

[6] W.J. Curran Jr., C.B. Scott, J. Horton, J.S. Nelson, A.S. Weinstein, A.J. Fischbach, C.H. Chang, M. Rotman, S.O. Asbell, R.E. Krisch, D.F. Nelson. *Recursive Partitioning Analysis of Prognostic Factors in Three Radiation Therapy Oncology Group Malignant Glioma Trials*. Journal of the National Cancer Institute, Vol. 85. No. 9, May 5, 1993. p.704-710.

[7] N. Alexeyeva, A. Alexeyev, E. Verbitskaya, E. Krupitsky. *The logical principle of projectivity and finite geometries in the statistical analysis of medical data.* Mathematical models. Theory and applications. Issue 8. Eds. prof. Chirkov M.K. Research Institution of Mathematic and Mechanic, Saint-Petersburg University. 2007. P. 3-14. (in Russian)