

Use of the Computer Statistics in Oncology

M. Shoikhedbrod*

Commented [k1]: Pls. mention affiliations of author.

Abstract

Currently, the use of computer statistics and computer statistical modeling in oncology for obtaining an accurate diagnosis, determination of the choice of treatment method and its correction in the process of ongoing treatment, prediction of the outcome of the disease, and evaluation of the effectiveness of the chosen treatment tactics is a decisive factor. The use of computer statistics, based on an adapted scientific and statistical package of the SSP in oncology, which is the basis for the use of computer statistical modeling of oncological processes, plays an important role in the effective treatment of cancer patients in clinical practice, since it permits, based on the creation of a computer medical and statistical model, to actively engage in the care of individuals undergoing cancer treatment, modify this statement. Effective participation in the treatment of cancer patients occurs due to the implementation of the tactics of individual planning of the examination of the patient, individualized prognosis, which determine the possibility of an individual approach to the observation and postoperative treatment of the patient according to the constructed medical-statistical model. This paper presents the results of computer-statistical processing of information from cancer patients, using the SSP package of scientific and statistical programs, which became the basis for the development of a computer statistical optimal interpolation model for accurately predicting of the timing of the appearance of metastases after surgery and of evaluating of the effectiveness of treatment of malignant neoplasms.

Keywords: Computer statistics, computer statistical modeling, computer optimal interpolation, estimation of malignant new formations treatment efficiency, precise prognosis of metastases appearance timing

INTRODUCTION

The use of computer statistics and computer statistical modeling in oncology for obtaining an accurate diagnosis, determination of the choice of treatment method and its correction in the process of ongoing treatment, prediction of the outcome of the disease, and evaluation of the effectiveness of the chosen treatment tactics is a decisive factor.

The use of computer statistics, based on an adapted scientific and statistical package of the SSP [1] in oncology, which is the basis for the use of computer statistical modeling of oncological processes, plays an important role in the effective treatment of cancer patients in clinical practice, since it

permits, based on the creation of a computer medical and statistical model, to actively participate in the treatment of cancer patients.

*Author for Correspondence

M. Shoikhedbrod
E-mail: michaelshoikhedbrod@bell.net

Research and Development Department, Electromagnetic Impulse Inc., 21 Four Winds Drive, Unit 12, North York, Ontario M3J 1K7, Canada

Received Date: May 28, 2023
Accepted Date: June 08, 2023
Published Date: June 17, 2023

Citation: M. Shoikhedbrod. Use of the Computer Statistics in Oncology. Research & Reviews: Journal of Statistics. 2023; 12(1): 1–21p.

The Scientific Subroutine Package (SSP) includes the following modules: data preprocessing: correlation and regression: factor analysis (eigen value analysis).

The first module of the package permits the preliminary processing of patient data, including the calculation of the sum, average, standard deviations, minimums and maximums of all or selected patient

numbers for all or selected symptoms under study; selection of numbers of patients who are within the specified boundaries for all or selected symptoms under study; selecting a subsequence from a matrix of patient numbers and symptoms under investigation; determination of the number of passed patients; tabulation of the number of patients for one or two investigated symptoms; printout of a histogram of the frequencies of the number of patients relative to the intervals of the studied symptoms [2].

The second module of the package is of the greatest interest, since it allows to calculate the existing pairwise correlation (relationship) between all the studied symptoms (correlation matrix) and, in certain cases, to recognize (predict) a specific relationship (regression) with a certain accuracy between a pair of studied symptoms or the studied symptom with multiple symptoms using the Basic Polynomial Regression Program (POLRG), Basic Multilinear Regression Program (MULTR), and Basic Stepwise Multiple Regression Program (STPRG) [3].

The third module of the package: factor analysis of data determines the values of the correlation matrix and, on its basis, forms a matrix of factors, each of which represents a latent (hidden) variable symptom, including a strongly correlated set of symptoms, related to the object of study, in this case, a cancerous tumor or the appearance of metastases cancerous tumor.

Factor analysis is a technique that condenses a large set of symptom variables linked to observations into a smaller set of independent influential variables known as factors. It combines symptom variables that exhibit strong correlations into a single factor. Variable symptoms from different factors are weakly correlated with each other.

Factor analysis classifies the features (symptom variables) that describe observations (patients). A factor is a latent (hidden) symptom variable constructed in such a way that a correlation between a set of symptom variables present can be explained.

The concept of factor analysis is to "compress" information.

Factor analysis allows us to solve two important tasks of the researcher: to describe the measurement of an object in a complex and at the same time compactly [4, 5].

Thus, all three modules allow the doctor-researcher to obtain complete statistical information on the selected by doctor or of all of the symptoms of a specific localization of oncological disease of patients or on the selected oncological patient at the stage of preliminary statistical processing; determine a pair relationship (correlation) between all the symptoms of the localization of a malignant tumor, chosen by the doctor, and if it is detected (the correlation is equal to one), try to determine, using polynomial regression, the polynomial mathematical expression of this pair relationship, using multilinear regression, determine a linear mathematical expression for the linear relationship of one selected symptom of a specific disease through other five dependent symptoms of this disease, and using stepwise multiple regression, determine a linear mathematical expression for the linear relationship of one selected symptom of a specific disease through five dependent symptoms of this disease step by step in the second module.

The third module, using factor analysis, allows a large number of symptoms of a particular oncological disease of patients to be reduced to a smaller number of independent influencing symptoms, called factors: the variables of which are highly correlated with each other, combined into one symptom-factor.

The use of the results of computer statistical processing of medical information, carried out with the help of an adapted scientific and statistical package of the SSP, in oncology becomes the basis for building a computer statistical model of a cancerous tumor or the appearance of metastases of this tumor.

This paper presents the results of computer-statistical processing of information of cancer patients, using the SSP package of scientific and statistical programs, which became the basis for the development of a computer statistical optimal interpolation model for accurately predicting of the timing of the appearance of metastases after surgery and of evaluating of the effectiveness of treatment of malignant neoplasms.

MATERIALS

The application of the developed computer statistical optimal interpolation model for accurate prediction of the timing of the appearance of metastases after surgery and evaluation of the effectiveness of treatment was carried out based on the results of statistical processing of medical information of patients with detected metastases of lung cancer, breast cancer in women and men, malignant testicular tumors on the basis of the All-Union Scientific Cancer Center USSR Academy of Medical Sciences (Moscow, Russia) and the Republican Clinical Oncological Dispensary (RCOD, Tajikistan).

METHODS

For the preparation of information, medical data of patients with cancer of different localizations with their subsequent statistical processing in the mode of the automated workplace of a doctor, the author developed in computer language C an original software system for creating and managing a daily database of tumors DBMS (database management system) *Medical Commander* [6].

The developed software system occupied a minimum of RAM in size, since the system read, recorded and processed information directly from magnetic media (floppy disks, hard drives, USB drivers, etc.).

In real time, the system uses at least 10 tumor documents (tumor passport) at the same time, each with information memory, limited by the capacity of the magnetic drive and only 100 symptoms x 100 patients of digital and coded data for statistical processing of each document by an adopted by author in computer language C, a scientific statistical programs package SSP, all modules of which are called by the developed system for statistical processing of the selected or the entire existing database of the tumor passport at the program level.

The interactive mode was provided by the system in such a way that typical data (questions) common to the oncologist, previously recorded using the system on magnetic media, asked the doctor on the computer screen what he should fill in his daily medical work from the outpatient card or the history of the oncological patient.

Figure 1 illustrates the operating modes of the system displayed on the computer screen.



Figure 1. The regimes of the work of the system Medical Commander.

Here, the "Add" operating mode for creating a tumor passport and adding information to the created tumor passport; "Edit" operating mode for editing the information, available in the tumor passport; "List" operating mode for printing the information, existing in the tumor passport, on the computer screen; "Conv." operating mode for converting the information, available in the tumor passport into a form, acceptable by the SSP package for statistical processing, and implementing statistical processing for the selected or all modules of the package, and, final. "Quit" operating mode to exit from the system.

All questions (symptoms or factors) of a qualitative nature were reduced to answers yes or no, i.e., 1 or 0, which permitted to speed up the input, mathematical analysis and prediction of these symptoms. Quantitative values were entered in the usual digital form.

The Figure 2 illustrates a common form of an electronic "medical history" of the patient with the second number, filled out by an oncologist, using the *Medical Commander* system.

Figure 3 shows a general list of patient numbers, observed by the oncologist, on a computer screen, using the *Medical Commander* system.

The attending oncologist at any stage of monitoring the patient or his treatment can see on the computer screen the patient's surname and the name of the symptom from a special reference book-codifier, which is also stored on magnetic media.

At each stage of the work of the *Medical Commander* system, the selection of both patients and symptoms was provided in an interactive mode.

Figure 4 shows how the oncologist selects the symptom of the time of occurrence of metastases among all available patients for further mathematical processing.

As a result, an oncologist can receive on a computer screen, upon a selective or complete request, complete or selective information about all patients or selected patients in a single.

As a result, an oncologist can receive on a computer screen, upon a selective or complete request, complete or selective information about all patients or selected patients in a single medical history form for doctors and, if necessary, print this information.

This stage of the system involved storing the database in the form of a matrix, in which the numbers of patients were conditionally located in the horizontal direction above the top row of the matrix, and the symptoms of the disease were conditionally located in front of the first column of the matrix in the vertical direction.

The screenshot displays the 'Medical Commander, Version 1.1' interface. At the top, it reads 'Copyright (C) 1998-9 by Michael Shoikhedbrod'. Below this, the word 'Printing' is centered. The main content area shows a form for a patient's medical history. The patient's name is 'Dan Lynch, Canada, Toronto, Ont.'. The time of metastases appearance is '1'. Life time is '23'. Sex is 'M'. Age is '28'. There are four numbered lines for 'Associated chronic diseases (to AIDS)'. There are two numbered lines for 'Notes'. A green box labeled 'Z Kit' is visible in the top right corner of the form area.

Figure 2. The common form of the electronic "medical history" of second number patient.



Figure 3. The common list of the numbers of patients, with which works the doctor-oncologist, using the system *Medical Commander* on the computer screen.

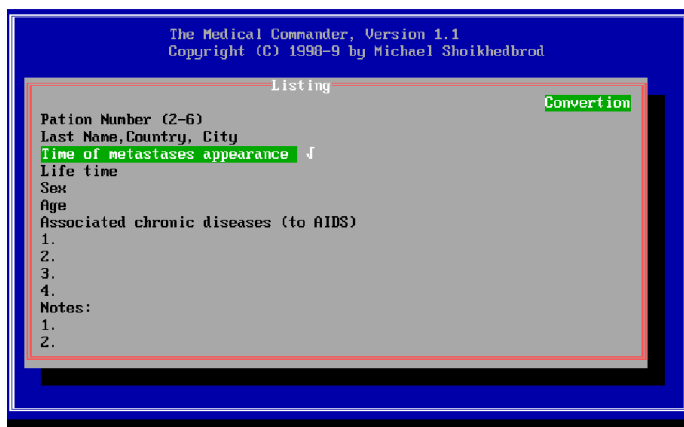
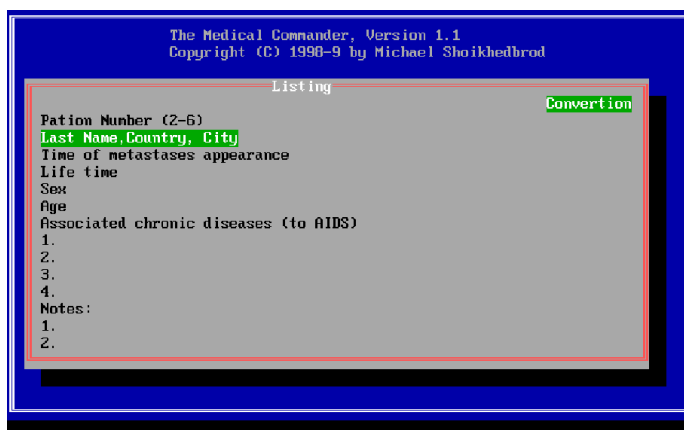


Figure 4. Selection of the symptom of the time of the appearance of metastases among all existing patients for further mathematical processing medical history form for doctors and, if necessary, print this information.

Thus, each value of the matrix element corresponded to a certain number of the patient and a certain symptom of the disease.

Statistical processing of the completed data base and analysis of the obtained data were carried out, using the IBM SSP (Scientific Subroutine Package) of scientific and statistical programs, adapted by the author in computer language C.

All source modules of the package were rewritten in the C programming language and compiled into batch files for a personal computer.

As noted earlier, the adapted package of scientific statistical programs of the SSP package was coordinated with the *Medical Commander* system.

The *Medical Commander* system, at the request of the researcher, called individual modules or all modules of the package for statistical processing and mathematical analysis of the data, contained in tumor documents for specific localizations at the program level.

The Scientific Subroutine Package (SSP) includes the following modules:

PRELIMINARY DATA PROCESSING

- Tally:** sum, average, standard deviations, the minimums and maximums;
- Bound:** the selection of observations, which are located in the assigned boundaries;
- Subst:** selection of subsequence from the matrix of observations;
- Absnt:** the determination of passed data;
- Tab 1:** the tabulation of data (one variable);
- Tab 2:** the tabulation of data (two variables);
- Submx:** the construction of the reduced matrix;
- Hist:** the printout of the histogram of frequencies relative to intervals;
- Plot:** the drawing of the separate interdependent variables relative to base variable.

CORRELATION AND REGRESSION

- Corre:** mean, standard (mean-square) deviations and correlation;
- Misr:** average, standard deviations, the third and fourth moments, correlation, the coefficients of simple regression and their standard errors, the determination of the possible omissions of data;
- Order:** regrouping cross-correlation;
- Multr:** multilinear regression;
- Minv:** for the matrix inversion of correlation coefficients;
- Gdata:** the formation of the matrix of data for the polynomial regression;
- Polrg:** the main program of polynomial regression;
- Stprg:** step-by-step multilinear regression;
- Probt:** probabilistic analysis;
- Canor:** canonical correlation.

FACTOR ANALYSIS (ANALYSIS OF EIGEN VALUES)

- Facto:** the main program of factor analysis;
- Corre:** mean, standard (mean-square) deviations and correlation;

Eigen: the calculation of eigen values and corresponding eigen vectors of correlation matrix;

Trace: the selection of the eigen values, which more or are equal to reference quantities indicated by the user;

Load: the calculation of the matrix of factors;

Varmx: the completing of the varimax rotation of the matrix of factors.

The first module of the package permits the preliminary processing of patient data, including the calculation of the sum, average, standard deviations, minimums and maximums of all or selected patient numbers for all or selected symptoms under study; selection of numbers of patients who are within the specified boundaries for all or selected symptoms under study; selecting a subsequence from a matrix of patient numbers and symptoms under investigation; determination of the number of passed patients; tabulation of the number of patients for one or two investigated symptoms; printout of a histogram of the frequencies of the number of patients relative to the intervals of the studied symptoms.

The second module of the package is of the greatest interest, since it allows to calculate the existing pairwise correlation (relationship) between all the studied symptoms (correlation matrix) and, in certain cases, to recognize (predict) a specific relationship (regression) with a certain accuracy between a pair of studied symptoms or the studied symptom with multiple symptoms using the Basic Polynomial Regression Program (POLRG), Basic Multilinear Regression Program (MULTR), and Basic Stepwise Multiple Regression Program (STPRG).

The third module of the package: factor analysis of data determines the values of the correlation matrix and, on its basis, forms a matrix of factors, each of which represents a latent (hidden) variable symptom, including a strongly correlated set of symptoms, related to the object of study, in this case, a cancerous tumor or the appearance of metastases cancerous tumor.

Variable symptoms from different factors are weakly correlated with each other.

Factor analysis classifies the features (symptom variables) that describe observations (patients). A factor is a latent (hidden) symptom variable constructed in such a way that a correlation between a set of symptom variables present can be explained.

The concept of factor analysis is to "compress" information.

Factor analysis allows us to solve two important tasks of the researcher: to describe the measurement of an object in a complex and at the same time compactly.

Thus, all three modules allow the doctor-researcher to obtain complete statistical information on the selected by doctor or of all of the symptoms of a specific localization of oncological disease of patients or on the selected oncological patient at the stage of preliminary statistical processing; determine a pair relationship (correlation) between all the symptoms of the localization of a malignant tumor, chosen by the doctor, and if it is detected (the correlation is equal to one), try to determine, using polynomial regression, the polynomial mathematical expression of this pair relationship, using multilinear regression, determine a linear mathematical expression for the linear relationship of one selected symptom of a specific disease through other five dependent symptoms of this disease, and using stepwise multiple regression, determine a linear mathematical expression for the linear relationship of one selected symptom of a specific disease through five dependent symptoms of this disease step by step in the second module.

The third module, using factor analysis, allows a large number of symptoms of a particular oncological disease of patients to be reduced to a smaller number of independent influencing

symptoms, called factors: the variables of which are highly correlated with each other, combined into one symptom-factor.

The results of main program of polynomial regression (POLRG) on a personal computer were as follows:

Data Input (Tables 1, 2)

- The total number of patients is 15.
- Independent variable among patients numbers:
 $X(i=1, \dots, 15) = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15.$
- Dependent symptom variable among patients:
 $Y(i=1, \dots, 15) = 10, 16, 20, 23, 25, 26, 30, 36, 48, 62, 78, 94, 107, 118, 127.$

Data Output (Table 3)

- The polynomial regression of order 1.
 The intersection $(b_0) -0.1387613E^{02}$
 The regression coefficients $0.8567849E^{01}$
- The polynomial regression of order 2.
 The intersection $(b_0) 0.1507547E^{02}$
 The regression coefficients $-0.1649572E^{01}, 0.6385619E^{00}$
- The polynomial regression of order 3.
 The intersection $(b_0) 0.1849422E^{02}$
 The regression coefficients $-0.3881073E^{01}, 0.9768364E^{00}, -0.1409456E^{-01}$
- The polynomial regression of order 4.
 The intersection $(b_0) -0.5267349E^{01}$
 The regression coefficients $0.1944943E^{02}, -0.5114683E^{01}, 0.5650751E^{00}, -0.1811317E^{-01}$

Table 1. Residual table 1.

S.N.	Value of X	Value of Y	Assessment of Y	Residual
1	1.00000	10.00000	9.61436	0.38564
2	2.00000	16.00000	17.40358	-1.40358
3	3.00000	20.00000	20.83865	-0.83865
4	4.00000	23.00000	22.22327	0.77673
5	5.00000	25.00000	23.42639	1.57361
6	6.00000	26.00000	25.88222	0.11778
7	7.00000	30.00000	30.59026	-0.59026
8	8.00000	36.00000	38.11533	-2.11533
9	9.00000	48.00000	48.58743	-0.58743
10	10.00000	62.00000	61.70206	0.29794
11	11.00000	78.00000	76.71997	1.28003
12	12.00000	94.00000	92.46704	1.53296
13	13.00000	107.00000	107.33398	0.33398
14	14.00000	118.00000	119.27710	1.27710
15	15.00000	127.00000	125.81982	1.18018

Table 2. Data input.

Patient No.	Variable Symptoms						Patient No.	Variable Symptoms					
	X1	X2	X3	X4	X5	X6		X1	X2	X3	X4	X5	X6
1	29	289	216	85	14	1	16	45	310	262	94	20	2
2	30	391	244	92	16	2	17	45	151	339	96	35	3
3	30	424	246	90	18	2	18	45	370	357	88	15	4
4	30	313	239	91	10	0	19	45	379	193	147	64	4
5	35	243	275	95	30	2	20	45	463	206	105	31	3
6	35	365	219	95	21	2	21	45	316	245	132	60	4
7	43	395	267	100	39	3	22	45	280	225	108	36	4
8	43	356	274	79	19	2	23	44	395	215	101	27	1
9	44	346	255	126	56	3	24	49	139	220	136	59	0
10	44	156	258	95	28	0	25	49	245	205	113	37	4
11	44	278	249	110	42	4	26	49	373	215	88	25	1
12	44	349	252	88	21	1	27	51	224	215	118	54	3
13	44	141	236	129	56	1	28	51	677	210	116	33	4
14	44	245	236	97	24	1	29	51	424	210	140	59	4
15	45	297	256	111	45	3	30	51	150	210	105	30	0

Commented [P2]: We have Change Column Head Some Table. please check if it is okay from your side.

Table 3. Data output: multiple linear regressions.

Independent symptom number X vs Y	Correlation	Regression coefficient
1	0.28422	0.01242
2	0.42189	0.00739
3	0.11900	0.01504
4	0.37822	0.00151
5	0.39412	0.04919
Dependent symptom number 6, Intersection -6.07928, Multiple correlation: 0. 73575		

Here, the value of X: the independent variable symptom; the value of Y: the dependent variable symptom; the assessment of Y: the dependent variable symptom calculated according to the found regression coefficients; the residual: the difference between the dependent variable symptom and the dependent variable symptom calculated according to the found regression coefficients.

The results of the main program of multilinear regression (MULTR) on a personal computer were as follows:

Tables 4 and 5 shows Residual Table 2 and Data Input Respectively

Here, the value of Y: the dependent variable symptom; the assessment of Y: the dependent variable symptom calculated according to the found regression coefficients; the residual: the difference between the dependent variable symptom and the dependent variable symptom calculated according to the found regression coefficients. In the case of stepwise multiple regression, the relationship between the dependent variable symptom (y) and the set of the independent symptoms (x_1, x_2, \dots, x_n) is established in order of importance of independent symptoms. The significance criterion is based on the reduction of the sum of squares, and the independent variable symptom most influencing on this decrease in this step, is inputted into the regression. As dependent variable symptom can be taken any variable from the source sets. Some set of variable symptoms can be forced inputted into the regression and some other set can to be dropped.

The results of the main program of stepwise multiple regressions (STPRG) on a personal computer were as follows:

Table 4. Residual table 2.

Patient No.	Value of Y	Assessment of Y	Residual
1	1.00000	0.48091	0.51909
2	2.00000	1.77670	0.22330
3	2.00000	2.14586	-0.14586
4	0.00000	0.82880	-0.82880
5	2.00000	1.90522	0.09478
6	2.00000	1.52125	0.47875
7	3.00000	3.46447	-0.46447
8	2.00000	2.25887	-0.25887
9	3.00000	3.80259	-0.80259
10	0.00000	1.02042	-1.02042
11	4.00000	2.49735	1.50265
12	1.00000	2.00066	-1.00066
13	1.00000	2.00735	-1.00735
14	1.00000	1.15308	-0.15308
15	3.00000	2.90446	0.09554
16	2.00000	1.83532	0.16468
17	3.00000	2.56004	0.43996
18	4.00000	3.45229	0.54771
19	4.00000	3.62661	0.37339
20	3.00000	2.68068	0.31932
21	4.00000	3.64885	0.35115
22	4.00000	1.86542	2.13458
23	1.00000	1.86542	2.13458
24	0.00000	1.86542	2.13458
25	4.00000	2.09863	-1.09863
26	1.00000	1.88027	-0.88027
27	3.00000	2.27646	0.72354
28	4.00000	4.51080	-0.51080
29	4.00000	3.95745	0.04255

Table 5. Data input.

Patient No.	Variable Symptoms						Patient No.	Variable Symptoms					
	X1	X2	X3	X4	X5	X6		X1	X2	X3	X4	X5	X6
1	29	289	216	85	14	1	16	45	310	262	94	20	2
2	30	391	244	92	16	2	17	45	151	339	96	35	3
3	30	424	246	90	18	2	18	45	370	357	88	15	4
4	30	313	239	91	10	0	19	45	379	193	147	64	4
5	35	243	275	95	30	2	20	45	463	206	105	31	3
6	35	365	219	95	21	2	21	45	316	245	132	60	4
7	43	395	267	100	39	3	22	45	280	225	108	36	4
8	43	356	274	79	19	2	23	44	395	215	101	27	1
9	44	346	255	126	56	3	24	49	139	220	136	59	0
10	44	156	258	95	28	0	25	49	245	205	113	37	4

11	44	278	249	110	42	4	26	49	373	215	88	25	1
12	44	349	252	88	21	1	27	51	224	215	118	54	3
13	44	141	236	129	56	1	28	51	677	210	116	33	4
14	44	245	236	97	24	1	29	51	424	210	140	59	4
15	45	297	256	111	45	3	30	51	150	210	105	30	0

Data Output (Tables 6–10)

Stepwise multiple regression

Number of patients 30

Number of variable symptoms 6

Dependent variable 6

Number forcedly inputted variable symptoms 0

Number of discontinued variable symptoms 0

Step 1

Inputted variable symptom 2

Amount of squares reduced at this step 10.300

Attitude reduced at this step 0.178

Accumulated amount of reduced square 10.300

Accumulated reduced relationship 0.178, 57.876

For the first inputted variable symptom

Multiple correlation coefficient 0.422

(Corrected on the degree of freedom) 0.422

Value F for dispersion analysis 6.063

Standard evaluation error 1.303

Step 2

Inputted variable symptom 5

Amount of squares reduced at this Step 13.3241

Attitude reduced at this step 0.230

Accumulated amount of reduced square 23.624

Accumulated reduced relationship 0.408, 57.867

For the second inputted variable symptom

Multiple correlation coefficient 0.639

(Corrected on the degree of freedom) 0.622

Value F for dispersion analysis 0.314

Standard Evaluation Error 1. 126

Step 3

Inputted variable symptom 3

Amount of squares reduced at this Step 7.572

Attitude reduced at this step 0.131

Accumulated amount of reduced square 31.106

Accumulated reduced relationship 0.539, 57.867

For the third inputted variable symptom

Multiple correlation coefficient 0.734

(corrected on the degree of freedom) 0.711
 Value F for dispersion analysis 10.137
 Standard evaluation error 1.013

Step 4

Inputted variable symptom 1
 Amount of squares reduced at this step 0.127
 Attitude reduced at this step 0.002
 Accumulated amount of reduced square 31.323
 Accumulated reduced relationship 0.541, 57.867
 For the fourth inputted variable symptom
 Multiple correlation coefficient 0.736
 (corrected on the degree of freedom) 0.699
 Value F for dispersion analysis 7.375
 Standard evaluation error 1.030
 (corrected on the degree of freedom) 1.088

Step 5

Inputted variable symptom 4
 Amount of squares reduced at this step 0.002
 Attitude reduced at this step 0.000
 Accumulated amount of reduced square 31.325
 Accumulated reduced relationship 0.541, 57.867
 For the fifth inputted variable symptom
 Multiple correlation coefficient 0.736
 (corrected on the degree of freedom) 0.684
 Value F for dispersion analysis 5.665
 Standard evaluation error 1.052
 (corrected on the degree of freedom) 1.133

Table 6. Data output (Step 1).

Variable Symptom Number	Regression coefficient	Standard Error of Regression coefficient
2	0.00521	0.00212
Intersection	0.62005	

Table 7. Data output (Step 2).

Variable Symptom Number	Regression coefficient	Standard Error of Regression coefficient
2	0.00632	0.00186
5	0.04316	0.01332
Intersection	-1.20349	

Table 8. Data output (Step 3).

Variable Symptom Number	Regression coefficient	Standard Error of Regression coefficient
2	0.00744	0.00172
5	0.05363	0.01258
3	0.01497	0.00551
Intersection	-5.53529	

Table 9. Data output (Step 4).

Variable Symptom Number	Regression coefficient	Standard Error of Regression coefficient
2	0.00741	0.00175
5	0.05076	0.01524
3	0.01493	0.00561
1	0.01226	0.03541
Intersection	-5.94617	

Table 10. Data output (Step 5).

Variable Symptom Number	Regression coefficient	Standard Error of Regression coefficient
2	0.00739	0.00186
5	0.04919	0.04141
3	0.01504	0.00635
1	0.01242	0.03635
4	0.00151	0.03679
Intersection	-6.07929	

Tables 11 and 12 shows residual Table 3 and data input, respectively. The results of the main program of factor analysis (FACTO) on a personal computer were as follows:

Table 11. Residual table 3.

Patient No.	Value of Y	Assessment of Y	Residual
1	1.00000	0.48090	0.51910
2	2.00000	1.77670	0.22330
3	2.00000	2.14586	-0.14586
4	0.00000	0.82880	-0.82880
5	2.00000	1.90522	0.09478
6	2.00000	1.52125	0.47875
7	3.00000	3.46447	-0.46447
8	2.00000	2.25887	-0.25887
9	3.00000	3.80259	-0.80259
10	0.00000	1.02042	-1.02042
11	4.00000	2.49735	1.50265
12	1.00000	2.00066	-1.00066
13	1.00000	2.00735	-1.00735
14	1.00000	1.15308	-0.15308
15	3.00000	2.90446	0.09554
16	2.00000	1.83532	0.16468
17	3.00000	2.56004	0.43996
18	4.00000	3.45229	0.54771
19	4.00000	3.62661	0.37339
20	3.00000	2.68068	0.31932
21	4.00000	3.64885	0.35115
22	4.00000	1.86542	2.13459
23	1.00000	2.09863	-1.09863
24	0.00000	1.97217	2.13458
25	4.00000	1.41254	2.58746
26	1.00000	1.88027	-0.88027

27	3.00000	2.27646	0.72354
28	4.00000	4.51080	-0.51080
29	4.00000	3.95746	0.04254

Table 12. Data input.

Patient No.	Variable symptoms								
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
1	7	7	9	7	15	36	60	15	24
2	13	18	25	15	13	35	61	18	30
3	9	18	24	23	12	43	62	14	31
4	7	13	25	36	11	12	63	26	32
5	6	8	20	7	15	46	18	28	15
6	10	12	30	11	10	42	27	12	17
7	7	6	11	7	15	35	60	20	25
8	16	19	25	16	13	30	64	20	30
9	9	22	26	24	13	40	66	15	32
10	8	15	26	30	13	10	66	25	34
11	8	10	26	24	17	40	20	30	18
12	9	12	28	11	8	45	30	15	19
13	11	17	21	30	10	45	60	17	30
14	9	16	26	27	14	31	59	19	17
15	10	15	24	18	12	29	48	18	26
16	11	11	30	19	19	26	57	20	30
17	16	9	16	20	18	31	60	21	17
18	9	8	19	14	16	33	67	9	19
19	7	18	22	9	15	37	62	11	20
20	8	11	23	18	9	36	61	22	24
21	6	6	27	23	7	40	55	24	31
22	10	9	26	26	10	37	57	27	29
23	8	10	26	15	11	42	59	20	28

Data Output (Tables 13–16)

Factor analysis
 Patients number 23
 Variable symptoms number 9

Table 13. Correlation coefficient matrix.

Line 1									
1.00000	0.34937	0.11975	0.12102	0.21917	-0.09549	0.20901	-0.12903	0.05818	
Line 2									
0.34987	1.00000	0.41311	0.35572	-0.08243	-0.09103	0.29622	-0.32044	0.35387	
Line 3									
0.11975	0.41311	1.00000	0.41512	-0.43179	-0.08346	-0.10252	0.03215	0.27833	
Line 4									
0.12102	0.35572	0.41512	1.00000	-0.31288	-0.50365	0.49856	0.22539	0.59890	
Line 5									
0.21917	-0.08243	-0.43179	-0.31288	1.00000	-0.23000	0.03310	-0.00475	-0.30341	

Line 6								
-0.0955	-0.09100	-0.08345	-0.50365	-0.23003	1.00000	-0.44520	-0.25441	-0.37456
Line 7								
0.20901	0.29622	-0.10252	0.49856	0.03310	-0.44520	1.00000	-0.28050	0.60124
Line 8								
-0.12903	-0.32041	0.03215	0.22539	-0.00475	-0.25441	-0.28050	1.00000	0.13516
Line 9								
0.05818	0.35387	0.27833	0.59390	0.30341	-0.37456	0.60124	0.13516	1.00000

Table 14. Data output eigen vectors.

Vector 1								
0.16437	0.34836	0.28797	0.49661	-0.16806	-0.32922	0.39935	0.01287	0.47518
Vector 2								
0.34837	0.06552	-0.44647	-0.11893	0.61210	-0.26428	0.38860	-0.24845	-0.06014
Vector 3								
-0.29899	-0.46825	-0.23534	0.17377	0.14468	-0.43545	0.01881	0.61587	0.12470
Vector 4								
0.54441	0.16909	0.38289	0.04163	0.30537	-0.16163	-0.43411	0.4028	3-0.23789

Table 15. Data output matrix of factors (4 factors).

Variable Symptom 1			
0.28232	0.44663	-0.37286	0.56203
Variable Symptom 2			
0.59831	0.08400	-0.58394	0.17457
Variable Symptom 3			
0.49460	-0.57240	-0.29348	0.39528
Variable Symptom 4			
0.85293	-0.15248	0.21671	0.04297
Variable Symptom 5			
-0.28865	0.78475	0.18043	0.31525
Variable Symptom 6			
-0.56544	-0.33882	-0.54304	-0.16686
Variable Symptom 7			
0.68590	0.49821	0.02345	-0.44816
Variable Symptom 8			
0.02211	-0.31853	0.76803	0.41587
Variable Symptom 9			
0.81614	-0.07710	0.15551	-0.24559

Table 16. Data output rotated matrix of factors (4 factors).

Variable Symptom 1			
0.05498	0.07183	-0.05578	0.85017
Variable Symptom 2			
0.29329	-0.39653	-0.35581	0.60550
Variable Symptom 3			
0.05114	-0.82494	0.15068	0.32984
Variable Symptom 4			

0.74041	-0.41401	0.24580	0.13972
Variable Symptom 5			
-0.09091	0.80662	0.13525	0.39228
Variable Symptom 6			
-0.68286	-0.21579	-0.44983	-0.20503
Variable Symptom 7			
0.86997	0.18299	-0.34918	0.08830
Variable Symptom 8			
0.03602	-0.05500	0.91376	-0.15962
Variable Symptom 9			
0.80532	-0.32759	0.00994	-0.02380

Eigen values

2.94989, 1.64368, 1.55515, 1.06579

Accumulated Eigen Value Relationships

0.32777, 0.51040, 0.68319, 0.80161

Table 17. Compatibility check.

Variable Symptom	Initial	Final	Difference
1	0.73409	0.73408	0.00001
2	0.73649	0.73648	0.00001
3	0.81464	0.81463	0.00001
4	0.79955	0.79955	0.00001
5	0.83109	0.83108	0.00001
6	0.75725	0.75724	0.00001
7	0.92007	0.92006	0.00001
8	0.86476	0.86475	0.00001
9	0.75652	0.75651	0.00001

Thus, the statistical processing of data of the documents of tumors on the concrete localizations, carried out by an oncologist based on an adapted scientific statistical package SSP, described in detail above, permits to conduct preliminary general statistical processing of data, to carry out correlation and regression analysis of the numerical data symptoms, and also to construct the correlation matrix of factors using the factor analysis (Table 17).

The results of statistical processing permit to oncologist to establish the laws of the distribution of different symptoms or factors among oncological patients with the favorable and unfavorable outcomes of treatment, and also to evaluate the functional connections between pair and group of the investigated symptoms and to determine the significances of symptoms or factors (value of correlation).

A similar analysis permits to make the specific clinical conclusions and to plan the tactics of oncological patient's management.

The use of the results of computer statistical processing of medical information, carried out with the help of an adapted scientific and statistical package of the SSP in oncology becomes the basis for building a computer statistical model of a cancerous tumor or the appearance of metastases of this tumor.

For predicting the progression of the tumor process, an interpolation method of mathematical modeling was used.

Interpolation in computational mathematics is a method of the task involves determining intermediate values of nodal points within a given discrete set of known values.

When performing scientific and engineering calculations, it is common to work with sets of values that have been obtained through experimentation or random selection. As a result, the following mathematical problem arises, which formulates the essence of the interpolation method. Let us assume that the interval is given by the grid $\omega = \{x_0 = a < x < \dots < x_n = b\}$ and at its nodal points the values of the function $y(x)$ are given equal to $y(x_0) = y_0, \dots, y(x_i) = y_i, \dots, y(x_n) = y_n$. It is necessary to construct an interpolant, a function $f(x)$, coinciding with the function $y(x)$ at the grid nodes:

$$f(x_i) = y_i, \quad i = 0, 1, \dots, n \quad (1)$$

The basic purpose of interpolation is to obtain the efficient algorithm for swiftly calculating the values of $f(x)$ for the values x that are not present in the data table.

What is the process of choosing the interpolant function $f(x)$ and how can we approximate the error between the actual value $y(x)$ and the interpolant $f(x)$

Typically, interpolating functions $f(x)$ are constructed by combining elementary functions in the form of linear combinations

$$f(x) = \sum_{k=0}^n c_k F_k(x) \quad (2)$$

$F_k(x)$ represents a set of predetermined linearly independent functions, while c_0, c_1, \dots, c_n represent coefficients that have not yet been determined.

Consequently, derive a system of $n+1$ equations that pertain to the coefficients $\{c_k\}$.

$$\sum_{k=0}^n c_k F_k(x(i)) \quad (3)$$

Suppose that the system of functions $F_k(x)$ is constructed in a way that ensures the determinant of the system is non-zero for any choice of nodal points $a = x_0 < x_1 < \dots < x_n = b$.

$$\begin{aligned} &F_0(x_0) F_1(x_0) F_2(x_0) \dots F_n(x_0) \\ \Delta F = &F_0(x_1) F_1(x_1) F_2(x_1) \dots F_n(x_1) \\ &F_0(x_n) F_1(x_n) F_2(x_n) \dots F_n(x_n) \end{aligned} \quad (4)$$

Then on assigned $y(i)$ ($i = 0, 1, \dots, n$) are uniquely determined the coefficients c_k ($k = 0, 1, \dots, n$).

According to authors [7, 8], an algorithm for optimal interpolation of medical information was developed. The algorithm involved modeling the calculation of interpolation points for medical data by comparing various interpolation methods in terms of the error $y(x)-f(x)$. The algorithm automatically selects the interpolation method with the smallest error value.

Paraphrased: After performing statistical processing, which includes running the histogram construction program (HIST) within the SSP package, the data are inputted. (Figure 5).

The entered data are utilized as the nodal points for performing optimal interpolation.

The calculations are performed sequentially to generate the interpolation points and the error $y(x)-f(x)$ using polynomial interpolation methods. The primary and interpolation functions are used to construct graphs based on the obtained data from the optimal interpolation. The numerical values of these functions are displayed on the computer screen (Figure 6).

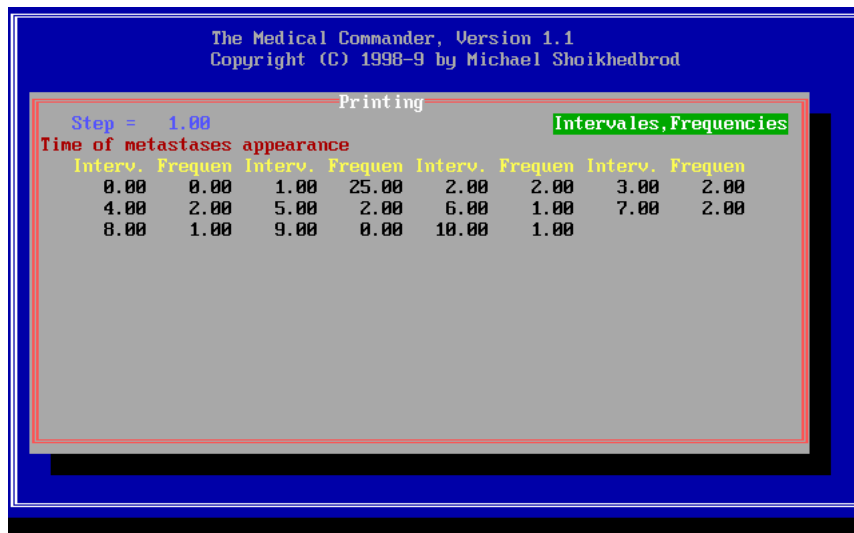


Figure 5. Following statistical processing, the input data consists of the frequencies of oncological patients.

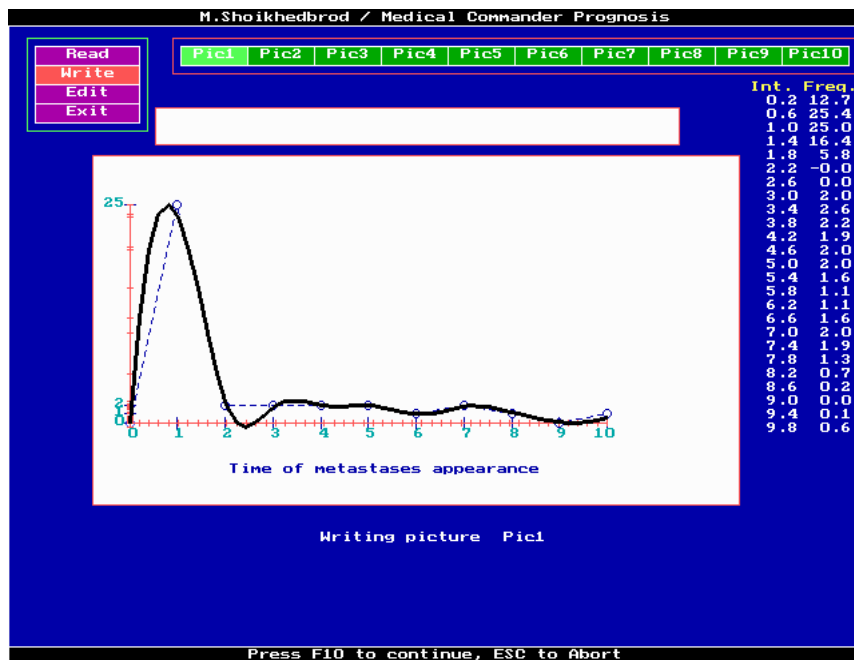


Figure 6. Using the data from optimal interpolation, graphs of the primary and interpolation functions are created, and the numerical values of these functions are displayed on the computer screen [9].

RESULTS AND DISCUSSION

Until now, oncologists believe [10] that the dependence of the timing of the appearance of tumor metastases among patients after surgery has an exponential form, i.e., the number of patients with metastases that appear after surgery decreases exponentially with time.

However, the authors [11, 12], who showed the first results of using computer optimal interpolation for predicting of the timing of the appearance of metastases among cancer patients after surgery, refuted this opinion.

For the first time, a specific period of time was shown when there is a surge (“hump”) in the number of patients with metastases that have appeared after a surgical operation.

The developed system of optimal interpolation was employed in clinical practice at the Republican Clinical Oncological Dispensary (RCOD, Tajikistan) to predict the progression of the tumor process in cancer patients

In this direction [12], patients with detected metastases of lung cancer, breast cancer among women and men, stomach and colon cancer, cancer cervix, malignant neoplasms of testicular tumors, lymphogranulomatosis among children were carried out.

Figure 7 shows that the dependence of the timing of the appearance of tumor metastases among patients after a surgical operation does not have an exponential form, i.e., the number of patients with tumor metastases that have appeared does not decrease exponentially in time, but a “hump” among patients with metastases after a surgical operation appears. The highest point of this “hump” occurred during the 21st month within the time frame of 12 to 30 months after the operation.

Therefore, it is highly likely that the occurrence of metastases among lung cancer patients will occur precisely during this particular period.

The residual peaks in Figure 7 can be categorized as minor statistical variations in the exponential decline of the number of oncological patients experiencing metastases and relapses.

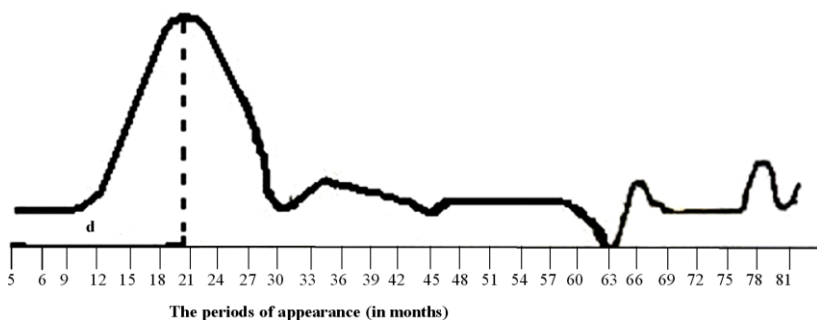


Figure 7. The progression of lung cancer in a group of patients who have undergone surgical operations. (d: the time when the peak number of patients with metastases occurs).

Commented [s3]: <Author> Please check: Figure 7 has been typed by us

As the 5-year period approaches, the frequency of examinations should be increased proportionally.

Patients who are predicted to have an unfavorable outcome should undergo monthly control examinations during the period from 12 to 30 months, with special attention given to the 21st month, when the appearance of metastases and relapse is anticipated.

Prognosis of the Progression of the Tumor Process of Cancer of Egg after the Conducted Treatment

The developed system of optimal interpolation, implemented on a personal computer, was utilized to forecast the progression of malignant tumors in the ovary and determine the appropriate timing for specialized anti-tumor therapy following surgical removal of the ovary. Thus, the conducted studies showed that the use of developed system of optimal interpolation on a personal computer on large groups of patients permitted to develop in clinical practice a new method for individual prediction of the timing of the appearance of metastases and a criterion that determines the effectiveness of the treatment of malignant neoplasms.

CONCLUSION

Thus, the results of the computer-statistical processing of information of oncological patients, using the package of scientific and statistical programs SSP, adopted by the author for use in the developed Medical Commander system, became the basis for the development of a computer-statistical optimal interpolation model for accurate prediction of the timing of the appearance of metastases after a surgical operation and evaluating of the effectiveness of treatment of malignant neoplasms.

The use of the developed system of optimal interpolation on a personal computer for accurate prediction of the timing of the appearance of metastases after a surgical operation among patients with lung cancer, breast cancer among women and men, stomach and colon cancer, cancer cervix, malignant neoplasms of testicular tumors, lymphogranulomatosis among children permitted to create a computer medical-statistical model of the appearance of metastases among patients after a surgical operation, numerically reflecting all the patterns of the course of the tumor process after conducted surgical operation.

The course of the tumor process after the surgical operation is accompanied by a surge of the appearance of the first metastases in the group of operated patients.

The developed system of optimal interpolation on a personal computer for accurate prediction of the timing of the appearance of metastases proved the possibility of using it to determine the effectiveness of the methods, used to treat cancer patients after a surgical operation.

An indicative result of the conducted research is the possibility of using of the developed system of optimal interpolation on a personal computer for accurate prediction of the timing of the appearance of metastases in wide medical practice.

On the basis of the obtained results, a methodology was developed that allows planning the optimal tactics of postoperative management of patients, choosing periods for follow-up examinations and regimens for preventive treatment of cancer patients.

The results of scientific research permitted not only to reduce the frequency, but also to significantly postpone the timing of the appearance of metastases after a surgical operation.

REFERENCES

1. IBM Corporation. System/360 Scientific Subroutine Package (360A-CM-03X) Version III. (1968).
2. Egoshin VL, Ivanov SV, Savvina NV, Kapanova GZ, Grjibovski AM. Descriptive statistics using R. *Ekolo Cheloveka/Hum Ecol.* 2018; (9): 55–64.
3. Bewick V, Cheek L, Ball J. Statistics review 7: Correlation and regression. *Crit Care.* 2003 Dec; 7: 1–9
4. Malinowski, Edmund R, Darryl G. Howery. Factor analysis in chemistry. New York: Wiley, 1980 Vol. 3.

5. Kline P. An easy guide to factor analysis, Routledge, 2014.
6. Shoikhedbrod MP. The use of developed computer medical process manager, “Medical Commander” for Application in Medical Practice, Int J Softw Comput Test. 2020; 6(2): 12–27.
7. Shoikhedbrod MP. Computer modeling and the new technologies in oncology. 2017; 1–116.
8. Shoikhedbrod MP. Computer modeling in physics and medicine, Lambert Academic publishing, Toronto, 2018
9. Shoikhedbrod DM. Programming of the Computer Manager of Medical Processes “Medic Commander” Software Application on Computer Language Turbo C. J Comput Sci Eng Softw Test. 2022; 8(1): 1–34. (e-ISSN: 2581-6969). Available from: <https://matjournals.co.in/index.php/JOCSES/article/view/25>
10. Enderling H, AJ Chaplain M. Mathematical modeling of tumor growth and treatment. Curr Pharma Design. 2014 Sep 1; 20(30): 4934–40.
11. Shoikhedbrod M. Principles of Interpolating Prognostication in Oncology. Research & Reviews: J Oncol Hematol. 2020; 9(2): 39–51. Available from: <https://medicaljournals.stmjournals.in/index.php/RRJoOH/article/view/2219>
12. Akhmedov BP, Akhmedova Sh. B. The results of the application of special mathematical methods of study on the computer in the study of prognostication and modeling of the processes of the generalization of malignant new formations. Tashkent. 2002.

Commented [k4]: Name & location of pub missing.