

**ACADEMIA DE ȘTIINȚE A MOLDOVEI
INSTITUTUL DE MATEMATICĂ ȘI INFORMATICĂ**

Cu titlu de manuscris
C.Z.U. 004.415.2

POPUKAYLO VLADIMIR

**SUPPORTUL DECIZIILOR ÎN BAZA EȘANTIOANELOR PASIVE
DE DIMENSIUNI MICI**

122.03 – MODELARE, METODE MATEMATICE, PRODUSE PROGRAM

Autoreferatul tezei de doctor în informatică

CHIȘINĂU, 2017

Lucrarea a fost realizată în laboratorul Sisteme de programare al Institutului de Matematică și Informatică al Academiei de Științe din Moldova

Conducător științific:

Cojocaru Svetlana, doctor habilitat în informatică, profesor, Institutul de Matematică și Informatică al Academiei de Științe din Moldova

Referenții oficiali:

1. Costăș Ilie, doctor habilitat în informatică, profesor, Academia de Studii Economice din Moldova.
2. Căpățână Gheorghe, doctor în științe tehnice, profesor, Universitatea de Stat din Moldova.

Componența Consiliului Științific Specializat:

1. Gaindric Constantin, președinte, doctor habilitat în informatică, membru-corespondent al AȘ din Moldova, profesor, Institutul de Matematică și Informatică a AȘ din Moldova.
2. Țițchiev Inga, secretar științific, doctor în informatică, conferențiar, Institutul de Matematică și Informatică AȘ din Moldova.
3. Mișcoi Gheorghe, doctor habilitat în științe fizico-matematice, academician al AȘ din Moldova, Institutul de Matematică și Informatică AȘ din Moldova și Universitatea Liberă Internațională din Moldova.
4. Cataranciuc Serghei, doctor habilitat în științe fizico-matematice, profesor, Universitatea de Stat din Moldova.
5. Palaghin Vladimir, doctor în științe tehnice, profesor, Universitatea Tehnică de Stat din Cercasc, Ucraina.
6. Corlat Andrei, doctor în științe fizico-matematice, conferențiar, Institutul de Matematică și Informatică al Academiei de Științe din Moldova.

Susținerea tezei va avea loc la 22.09.2017, la orele 15:00 , în ședința Consiliului Științific Specializat D 01.122.03-04 pe lângă Institutul de Matematică și Informatică al Academiei de Științe a Moldovei, str. Academiei 5, or. Chișinău, MD-2028, Republica Moldova, aud. 340.

Teza și autoreferatul pot fi consultate la Biblioteca Științifică Centrală AȘM „Andrei Lupan” precum și pe pagina web a C.N.A.A. (www.cnaa.md).

Autoreferatul a fost expediat la 12.07.2017.

Secretarul științific al Consiliului Științific Specializat

Țițchiev Inga, doctor în informatică, conferențiar

Conducător științific

Cojocaru Svetlana, doctor habilitat în informatică, profesor

Autor

Popukaylo Vladimir

© Popukaylo Vladimir, 2017

CARACTERIZAREA GENERALĂ A LUCRĂRII

Actualitatea cercetării. În lumea modernă o dezvoltare largă o au sistemele suport de luare a deciziilor (SSD) – sisteme computerizate automatizate interactive, care ajută persoana ce ia decizii (decidentul) în analiza diferitelor probleme dificile din punct de vedere al formalizării.

În același timp soluțiile universale sunt orientate, de regulă, spre matematicieni și ingineri, ceea ce îngreunează utilizarea lor de către specialiști din diferite alte domenii. În legătură cu aceasta o răspândire largă au obținut SSD orientate spre probleme, care se elaborează pentru o problemă concretă și se ajustează conform caracteristicilor individuale ale fiecărei situații specifice.

În analiza sistemelor complexe se aplică abordarea probabilistică și statistică în baza anumitor metode selective, care extind posibilitățile metodelor analitice.

Cu toate acestea în practică există astfel de domenii, în care caracteristicile obiectelor cercetate necesită analiza unui volum relativ mic de date experimentale, procesarea cărora prin metode probabilistice și statistice clasice deseori nu conduce la rezultate acceptabile.

Deși problema procesării statistice a eșantioanelor de volum mic este de aceeași natură, în diferite domenii ea se rezolvă prin diferite metode, ținând cont de specificul sistemelor studiate. În afară de aceasta, rămâne nerezolvată problema analizei multidimensionale a tabelelor experimentului pasiv. Astfel, devine actuală problema elaborării unor abordări în asistarea luării deciziilor în baza unor eșantioane pasive de volum mic, care ar putea fi utilizate pe larg pentru diferite clase de probleme.

Scopul și obiectele cercetării.

Obiectul cercetării îl constituie eșantioanele pasive de volum mic, obținute în diferite condiții.

Scopul cercetării. Elaborarea și argumentarea unor metode și tehnici specifice, utilizate în sisteme suport de luare a deciziilor la procesarea statistică a eșantioanelor pasive de volum mic, care permit obținerea unei informații mai precise și mai obiective cu privire la procesul examinat.

Pentru realizarea scopului enunțat au fost formulate următoarele **probleme de cercetare:**

1) realizarea analizei diferitor abordări și metode de identificare a datelor aberante în eșantioanele unidimensionale de volum mic;

2) cercetarea posibilităților utilizării metodei distribuției punctuale pentru procesarea eșantioanelor de volum mic;

3) elaborarea metodicii de aplicare a posibilităților metodei distribuției punctuale pentru modelarea multidimensională cu date pasive de volum mic;

4) utilizând metodica propusă:

- să se elaboreze modele matematice adecvate;
- să se aprecieze cantitativ calitatea modelelor elaborate.

5) să se analizeze capacitatea de determinare corectă a coeficienților de corelare a nivelului legăturii liniare de corelare, în cazul unui număr mic de observări;

6) să se verifice metodica elaborată pe date obținute în diferite domenii de cercetare;

7) să se elaboreze algoritmi și produse-program care realizează metodele laborioase;

8) să se elaboreze sisteme suport de luare a deciziilor pentru un șir concret de aplicații, în care obținerea datelor de volum mare nu este posibilă.

Baza teoretico-metodologică a cercetării o constituie lucrările consacrate analizei sistemice, sistemelor suport pentru decizii, statisticii matematice și modelării matematice. Obiectivele cercetării au fost elaborate în baza lucrărilor lui Gascarov D.V., Șapovalov V.S., Dolgov Iu. A., Stolearenco Iu. A., Crivenco M.P., Bolișev L.N., Șor Ia. B., Cobzari A. I., Orlov A. I., Pearson E.S., Efron B.M., Fisher R.A., de Winter J.C.F. și altora.

Metodele de cercetare: analiza literaturii științifice și metodice și experienței practice, sistematizarea rezultatelor obținute anterior pe tematica cercetării, compararea abordărilor existente la soluționarea problemelor, metodele teoriei luării deciziilor, analiza sistemică, teoria probabilităților și statistica matematică, modelarea, sinteza abordărilor și metodelor probabilistice și statistice existente.

Noutatea științifică a cercetării:

1) Au fost identificate criteriile sigure de determinare a datelor aberante și stabilite posibilitățile lor la determinarea măsurării anormale în eșantioane de volum mic, a fost propusă metodica determinării datelor aberante în baza sintezei abordărilor existente.

2) Au fost elaborate recomandări pentru efectuarea analizei liniare corelaționale a eșantioanelor multidimensionale de dimensiune mică.

3) S-a elaborat metodica construirii modelelor matematice bazată pe algoritmul măririi virtuale a eșantionului multidimensional folosind tabelele bidimensionale a densității de distribuție obținute prin metoda distribuției punctuale.

4) S-au propus recomandări de elaborare a sistemelor suport pentru decizii în baza datelor de volum mic. În baza lor s-au elaborat SSD pentru un șir de domenii.

Problemă științifică importantă, soluționată în această lucrare: determinarea metodicilor probabilistic - statistice de suport în luarea deciziilor în baza eșantioanelor de volum mic, obținute în cadrul experimentului pasiv.

Importanța teoretică a cercetării constă în posibilitatea îmbunătățirii calității deciziilor aprobate în baza abordărilor și algoritmilor propuși, bazate pe metode probabilistice și statistice de procesare a informației.

Importanța practică a rezultatelor lucrării. Metodicile de suport în luarea deciziilor, propuse în teză, pot fi aplicate pe larg la analiza statistică în diferite domenii de cercetare.

Valoarea rezultatelor rezumă în următoarele elemente:

1) sunt selectate criteriile, care cel mai precis determină măsurarea anomală posibilă în eșantionul unidimensional de volum mic;

2) sunt elaborate recomandările de realizare a analizei liniare corelaționale a tabelelor experimentului pasiv, în cazul volumului mic de date inițiale;

3) este elaborată și testată metodică construirea modelelor matematice, bazată pe mărirea virtuală a volumului eșantionului multidimensional de dimensiune mică, care permite construirea modelelor matematice, inclusiv pentru planuri suprasaturate;

4) s-au elaborat algoritmi și produse program, care realizează aceste metodici;

5) au fost elaborate SSD pentru domenii concrete, în care este imposibilă obținerea datelor de volum mare.

Veridicitatea și argumentarea rezultatelor lucrării este asigurată prin utilizarea metodelor de cercetare adecvate scopurilor și problemelor examinate, reprezentativitatea datelor experimentale, corectitudinea procesării lor cu aplicarea metodelor statisticii matematice.

Implementarea rezultatelor lucrării. Rezultatele lucrării sunt implementate în procesul didactic al catedrei „Biologia și fiziologia omului” și „Terapia Nr. 2” a facultății de medicină din cadrul Universității T.G. Șevcenco (or. Tiraspol), în activitatea laboratorului științific de cercetare „Modelarea matematică” a Universității T.G. Șevcenco (or. Tiraspol), în programul didactic de masterat pe profilul „Informatica și tehnica de calcul” al aceleiași universități.

Aprobarea rezultatelor lucrării. Rezultatele cercetării au fost prezentate la ședința catedrei „Tehnologii informaționale și automatizarea controlului proceselor de producere” a Universității T.G. Șevcenco (or. Tiraspol), seminarul laboratoarelor „Sisteme de programare” și „Sisteme informaționale” ale Institutului de Matematică și Informatică al Academiei de Științe a Republicii Moldova, precum și la următoarele conferințe:

– Conferința VIII-a internațională «Modelarea matematică în educație, știință și industrie», Tiraspol, 2013;

– 5th International Conference ICTEI-2015, Chișinău, 2015;

– Conferință a 23-a de Matematică Aplicată și Industrială CAIM 2015– CAIM, ROMANIA, Suceava, 2015;

- A IX Conferință internațională «Modelarea matematică în educație, știință și industrie», Tiraspol, 2015;
- Conferința științifico-practico din întreaga Rusie «Sisteme și tehnologii informaționale și în telecomunicații », FR, Kemerovo, 2015;
- Conferința științifică internațională a doctoranzilor «Tendințe ale dezvoltării științei moderne: viziunile tinerilor cercetători», Chișinău, 2016.
- The Fourth Conference of Mathematical Society of the Republic of Moldova dedicated to the centenary of Vladimir Andrunachievici (1917-1997), Chișinău, 2017.

Rezultatele științifice principale înaintate spre susținere:

- 1) Metodica detectării valorilor anormale în eșantioane de volum mic, care permite efectuarea celei mai precise proceduri de analiză a datelor pasive la prezența datelor aberante.
- 2) Metodica bazată pe mărirea virtuală a dimensiunii eșantionului prin metoda distribuției punctuale, care permite construirea modelelor matematice adecvate cu eșantioane pasive de volum mic.
- 3) Metodica detectării mărimii legăturii liniare corelaționale după eșantioanele de volum mic, care permite determinarea obiectivă a legăturii puternice de corelare nu mai mică de valoarea coeficientului de corelare $r=0,8$; procedurile măririi virtuale statistice a informației («*bootstrapping*») reduc precizia analizei de corelare efectuate, măbind dispersarea valorilor obținute.

La tema tezei au fost publicate 16 lucrări.

Structura și conținutul tezei. Lucrarea constă din introducere, trei capitole, concluzii și recomandări, bibliografie din 160 titluri, 4 anexe, 23 tabele, 30 figuri.

Cuvinte-cheie: teoria luării deciziilor, sisteme pentru suport decizii, eșantioane mici, experiment pasiv, analiză corelațională, analiză regresivă, date aberante, metoda distribuțiilor punctuale.

Conținutul tezei. Metodele clasice probabilistice și statistice de suport în luarea deciziilor nu întotdeauna dau rezultate acceptabile la procesarea datelor de volum mic. De aceea, scopul cercetării îl constituie elaborarea și argumentarea metodelor și tehnicilor specifice de procesare statistică a eșantioanelor de volum mic, care vor permite obținerea unei informații mai obiective pentru adoptarea deciziilor.

În introducere este argumentată actualitatea temei, determinat obiectul cercetării, formulate scopurile și sarcinile cercetării, determinată baza teoretico-metodologică a lucrării de disertație, determinate metodele de cercetare, formulată noutatea științifică, importanța teoretică și practică a

cercetării, formulate principiile de bază care sunt înaintate spre susținere, prezentate informațiile despre aprobarea și implementarea rezultatelor.

În primul capitol «Bazele metodologice și teoretice ale suportului luării deciziilor în condițiile eșantionului de volum mic » este examinată problema analizei sistemelor complexe și construirii SSD.

În prima secțiune sunt cercetate metodele și posibilitățile luării deciziilor, formulate definițiile SSD, descrise problemele, ce pot fi soluționate de către asemenea sisteme, se propune clasificarea lor. Este menționat, că SSD orientate spre probleme sunt pe larg răspândite în diferite domenii de cercetare, precum: medicina, industria, pedagogia, biologia și altele. Este prezentată schema adoptării deciziei [25], care constă din următoarele etape: descrierea domeniului ales, analiza situației, formularea problemei, analiza informației disponibile, elaborarea variantelor de soluționare, aprecierea aplicabilității variantelor de soluționare și consecințele lor, alegerea variantei optime a soluției, precum și verificarea și emiterea deciziei. Schema prezentată a luării deciziei nu reflectă complexitatea întregului proces. În practică, acest proces poate fi iterativ, de exemplu, evaluarea aplicabilității variantelor de soluții și consecințelor lor poate fi efectuată de câteva ori, până la momentul când riscurile de la luarea deciziilor vor fi minimale. De asemenea, schema admite executarea în paralel a procedurilor și operațiunilor separate, cum ar fi, spre exemplu, determinarea tipului de date ce descriu procesul cercetat și a metodelor de procesare a lor. Eficiența acestui proces în mare parte depinde de metodele utilizate la executarea fiecărei din etapele indicate ale schemei procesului de luare a deciziilor. Este efectuată o clasificare a metodelor de luare a deciziilor. Se examinează deosebirile metodelor formale și neformale, se descriu etapele luării deciziilor cu utilizarea abordării probabilistico-statistice.

În secțiunea a doua sunt descrise dificultățile aplicării metodelor probabilistice și statistice clasice de luare a deciziei în condițiile eșantionului de volum mic, se dovedește actualitatea problemei despre volumul necesar al eșantionului pentru diferite cercetări; se aduc exemple de domenii de cercetare, în care caracteristicile obiectelor nu permit obținerea unor volume mari de date experimentale. În astfel de cazuri decidentul trebuie să ia decizia în baza datelor obținute de dimensiuni mici, ceea ce conduce la apariția următoarelor probleme răspândite la evaluarea lor cu ajutorul metodelor probabilistice și statistice:

- odată cu micșorarea dimensiunii eșantionului se măresc intervalele de încredere și erorile probabilistice [28];

- în condițiile eșantionului de volum mic dispersia ansamblului de eșantioane nu poate fi utilizată în calitate de evaluare a dispersiei generale [3];

- în conformitate cu teorema centrală de limită [26] presupunerea despre distribuția normală a tuturor valorilor posibile este justă doar pentru un volum considerabil al eșantionului;

- abordările clasice la evaluarea legilor de distribuție a eșantioanelor, atât parametrice, cum ar fi criteriul Pearson și *RS*-criteriul, cât și ale celor neparametrice, de exemplu, criteriile de tip Kolmogorov-Smirnov și *W*-criteriul lui Shapiro-Wilk, nu permit să se judece cu o precizie suficientă despre apartenența ansamblului general la legea concretă de distribuție [5, 6];

- rezultatele procesării datelor, obținute la utilizarea diferitelor criterii, cum ar fi, de exemplu, Cochran și Grubbs, sunt dependente de dimensiunile eșantioanelor studiate [8];

- evaluarea parametrilor necunoscuți, realizată pe calea maximizării funcției apropierei de adevăr, care poate fi puternic distorsionată pentru cazul eșantioanelor de volum mic;

- aplicarea cuantilelor de distribuție Student Fischer presupune adoptarea ipotezei despre normalitatea legii de distribuție [4], în același timp erorile pentru majoritatea măsurărilor fizice sunt diferite de cele normale [10, 11];

- aplicarea metodelor Monte-Carlo se confruntă cu imperfecțiunea generatoarelor de numere pseudoaleatoare, precum și cu multitudinea infinită a devierilor posibile ale valorii aleatoare de la familia parametrică de distribuții [2, 9].

În afară de aceasta, se indică problemele de bază a analizei eșantioanelor multidimensionale de volum mic, care conduc la faptul, că odată cu creșterea volumului datelor, de regulă se reduce calitatea metodelor aplicate de analiză a datelor.

În secțiunea trei este examinată problema definirii termenului „eșantion mic” și definirea limitelor eșantioanelor mici și medii. Gaskarov și Șapovalov în lucrarea sa «Eșantion de volum mic» [1] scriu că eșantionul poate fi considerat drept unul mic, dacă la procesarea lui cu metode bazate pe gruparea observărilor nu este posibil de atins precizia și veridicitatea stabilită. Prin urmare, la procesarea eșantioanelor de volum mic trebuie să renunțăm la aceste metode și să utilizăm informația despre fiecare realizare aparte a eșantionului. Profesorul A. I. Orlov în lucrarea [9] afirmă că sub „eșantion de volum mic” se subînțelege un asemenea eșantion, pentru care nu pot fi aplicate rezultatele bazate pe teoremele limită. Astfel, în fiecare problemă concretă apare necesitatea de a separa volumurile finale ale eșantionului în două clase – cele pentru care pot fi aplicate teoremele limită, și cele pentru care acest lucru nu poate fi făcut din cauza riscului obținerii unor rezultate eronate. Este menționat că diferiți cercetători adoptă limita de sus a eșantionului mic în diapazonul de la 10 până la 30, fără a se argumenta prin careva date numerice. Prin urmare, problema de determinare a limitelor între eșantioanele de volum mic și mediu rămâne a fi una actuală. Analiza literaturii la această temă a arătat că una din puținele abordări argumentate la soluționarea acestei probleme este cercetarea [23], care s-a efectuat cu ajutorul experimentului

statistic de simulare în baza metodei *bootstrap*, în care s-a calculat limita de jos a evaluării intervalului celui mai probabil coeficient de corelare. În a doua cercetare la această temă [27] s-a efectuat analiza volumelor eşantioanelor, bazată pe utilizarea criteriului Pearson. Astfel, bazându-ne pe rezultatele a două cercetări independente se poate stabili că limita de sus a eşantionului de dimensiune mică este valoarea $n=20$.

În secțiunea patru este prezentat un rezumat succint al metodelor existente specializate în procesarea eşantioanelor de volum mic, bazate pe folosirea fiecărei realizări aparte, ce permit îmbunătățirea indicilor la evaluarea densității distribuției valorii aleatorii, cum ar fi:

1. Metoda depozitelor dreptunghiulare;
2. Metoda reducerii incertitudinii;
3. Metoda comprimării ariei existenței legilor integrale de distribuție;
4. Metoda medianelor consecutive;
5. Metoda adaosurilor (complementării) de simulare;
6. Metoda distribuției punctuale.

Din analiza efectuată se observă că majoritatea acestor metode se bazează pe utilizarea informației despre fiecare realizare în parte și reprezintă o dezvoltare a metodei depozitelor dreptunghiulare. De asemenea este prezentat algoritmul complet al distribuțiilor punctuale [23], în care în calitate de informație adițională se utilizează cunoștințele despre tipul legii de distribuție a valorii aleatoare. Sunt indicate intervalele măsurării valorii controlate X și formulele de aproximare pentru calcularea coeficienților auxiliari prin metoda distribuției punctuale pentru legea normală și exponențială, precum și pentru legea distribuției Weibull.

Secțiunea cinci este dedicată concluziilor și propunerilor conform rezultatelor capitolului. În baza materialului expus în capitol se concludă că în momentul de față este actuală problema elaborării SSD orientate spre probleme, destinate procesării datelor multidimensionale de volum mic. Pentru atingerea acestui scop este necesar:

1. De elaborat o metodică specializată de asistare în luarea deciziilor pentru cazurile procesării datelor multidimensionale de volum mic, bazată pe abordarea probabilistico-statistică, precum și pe analiza, sinteza și dezvoltarea metodelor existente.
2. De determinat metodele optime de luare a deciziilor referitor la identificarea și eliminarea erorilor grave în eşantioanele de volum mic.
3. De elaborat metodologia construirii tabelului virtual multidimensional de date conform eşantioanelor de volum mic în baza metodei distribuției punctuale.
4. De analizat și identificat metodica optimală de determinare a valorii legăturii liniare de corelare conform eşantioanelor de volum mic.

5. De analizat particularitățile metodelor modelării matematice cu date pasive în cazul eșantioanelor de volum mic.

În al doilea capitol «Elaborarea metodicilor de luare a deciziilor la procesarea datelor de volum mic» se examinează aplicabilitatea diferitor metode probabilistice și statistice pentru procesarea datelor de volum mic.

În prima secțiune se discută problema identificării datelor aberante în eșantioanele de volum mic. Această procedură este una obligatorie pentru procesarea primară a tabelelor experimentului pasiv, deoarece prezența unei singure măsurări anormale poate conduce la erori mari în evaluarea momentelor eșantioanelor și verificarea diferitelor ipoteze statistice. A fost realizată analiza metodicilor existente de identificare a datelor aberante în eșantioanele unidimensionale, și anume: metoda Irwin, criteriul Student, criteriul celei mai mari devieri absolute, criteriul devierii relativ maximale, criteriul Romanovski, metoda intervalului de variație, criteriul 3 Sigma, criteriul Right, criteriul Grubbs, , Q -criteriul (Dixon), criteriul Lvovski, criteriul Chauvin, criteriul David, criteriul Hoaglin-Iglewicz, L -criteriul (Criteriul Tietjen-Moore), criteriul Smolyak-Titarenko, criteriul Brodskii-Bațan-Vlasenko, criteriul Kimber. În baza analizei efectuate [12, 16] se propune utilizarea pentru aceste scopuri a criteriului Lvovski (în cazul legii normale de distribuție), Dixon (pentru diverse legi de distribuție), Irwin (pentru $n \geq 10$) sau Tietjen-Moore (având cea mai mică cantitate de erori de tip I), care asigură cea mai mare precizie pentru identificarea măsurărilor anormale în eșantioane de volum mic. De asemenea se studiază influența datelor aberante pereche și identificarea lor în tabelul experimentului pasiv.

În secțiunea a doua se analizează posibilitățile diferitelor abordări pentru aprecierea mărimii legăturii corelaționale liniare între două valori în cazul numărului mic de observații. Se examinează abordarea clasică de determinare a mărimii legăturii corelaționale liniare prin criteriul Pearson, aplicarea la *bootstrapping*, ca metodă de randomizare și mărire virtuală a informației procesate, precum și aplicarea metodei Cebâșev la mărirea volumului de informație prin metoda distribuției punctuale. Se arată că niciuna din aceste metode nu dispune de precizie suficientă pentru aplicarea analizei clasice corelaționale în cazul eșantioanelor de volum mic [14]. În particular, aplicarea metodei *bootstrap* mărește dispersia valorilor coeficientului calculat de corelare Pearson, în comparație cu metoda clasică. Iar aplicarea metodelor de distribuție punctuală și metodei Cebâșev conduc la imposibilitatea determinării nivelului înalt de dependență liniară între valori, atât cu ajutorul criteriului Pearson, cât și cu ajutorul indicelui modificat Fehner. La determinarea factorilor puternic legați se recomandă să nu fie folosite procedurile de randomizare și să fie aplicat pragul coeficientului $r=0,75$, deoarece valorile mai mici decât acest coeficient nu pot indica existența unei

legături corelaționale liniare puternice cu precizia necesară. De asemenea este prezentat algoritmul determinării legăturilor principale în matricea corelațională prin metoda pleiadelor corelaționale.

În secțiunea a treia se analizează metodică construirea tabelului extins virtual al experimentului pasiv în baza valorilor densității probabilităților, obținute din tabelele distribuției virtuale bidimensionale prin metoda distribuției pe puncte. La efectuarea transformărilor se construiesc tabele virtuale de date, pentru fiecare rând l al tabelului datelor inițiale, în care se introduc coloanele X_{ij} și X_{il} din tabelele corespunzătoare ale densității nenormate a probabilităților. Analiza corelațională și regresivă pentru astfel de tabele nu este posibilă din cauza unui număr mare de lacune în valori, în legătură cu aceasta rândurile incomplete sunt șterse din analiză, iar tabelele rămase se unesc în ordinea numerotării rândurilor datelor inițiale. De asemenea este prezentat exemplul numeric al calculului algoritmului descris [29] menționându-se uniformitatea tabelului de date inițiale și celor obținute, în baza imperceptibilității mediilor aritmetice și dispersiei eșantioanelor pentru fiecare din factorii și valorile emise după criteriul Student și Fischer. Se argumentează metodică identificării coeficientului de corelare după mărirea virtuală a informației prin metoda multidimensională de distribuție punctuală. Se arată că aplicarea unei asemenea abordări permite îngustarea intervalului de încredere la aprecierea coeficientului de corelare Pearson după datele de dimensiune mică, precum și evitarea determinării eronate a legăturii negative de corelare [13]. În cazul detectării în eșantioane $n > 10$ a legăturilor liniare de corelare $r > 0,75$, de asemenea trebuie de admis această legătură observată drept una puternică și semnificativă. În datele cu volum de $10 \geq n \geq 7$ legătura observată trebuie admisă ca una semnificativă pentru $r \geq 0,8$. La volumul eșantionului cercetat $n \leq 6$, valoarea coeficientului de prag al corelării trebuie admis la nivelul $r = 0,9$. Se concludă că aplicarea mării virtuale a tabelului de date inițiale prin metoda propusă, spre deosebire de metodele *bootstrapping*, permite păstrarea cunoștințelor despre tipul legii de distribuție a valorii aleatorii și despre mărirea legăturii liniare de corelare între factorii cercetați, ce va permite luarea unor decizii mai argumentate cu o cantitate mică de date inițiale, precum și să îmbunătățească calitatea aprecierii statistice prin metode parametrice.

În secțiunea patru se examinează problemele evaluării calității deciziilor luate în baza abordării probabilistico-statistice; se formulează conceptul gradului de adecvare a modelului matematic, sub care se subînțelege: 1) descrierea corectă și calitativă a calităților evaluate ale obiectului (procesului): de exemplu, a posibilității în baza cercetării de a face concluzii corecte despre direcția schimbărilor anumitor caracteristici cantitative ale acestor proprietăți, despre legătura lor reciprocă, despre caracterul fluctuațiilor obiectului sau procesului, stabilitatea stării sale sau evoluției și altele, iar în baza acestor date de a lua o decizie corectă; 2) descrierea corectă cantitativă a proprietăților evaluate cu o anumită precizie rațională.

În compartimentul respectiv se compară calitatea modelelor matematice, construite în diferite moduri. Se conchide că aplicarea metodicii elaborate, într-un șir de cazuri ajută la obținerea unor rezultate mai calitative de modelare matematică, ce poate să permită cercetătorului să obțină decizii mai eficiente la analiza datelor multidimensionale de volum mic.

Secțiunea a cincea este dedicată analizei rezultatelor cercetării. Se accentuează că a fost cercetată procedura căutării datelor aberante în eșantioane de volum mic și stabilite criteriile care soluționează în cel mai reușit mod această problemă, au fost propuse recomandări de realizare a procedurii de înlăturare a datelor aberante; cercetate posibilitățile abordărilor existente în identificarea legăturii liniare de corelare în eșantioane de volum mic, conform rezultatelor lucrării s-a propus procedura analizei liniare corelaționale pentru problemele identificării factorilor strâns legați și reducerea ulterioară a spațiului factorial, fără mari pierderi de informație; elaborată metoda de construire a eșantionului virtual multidimensional, folosind metoda distribuției punctuale și conectarea datelor după valorile maxime ale densităților probabilităților; arătat că aplicarea acestei metode permite obținerea unui tabel de peste 10 ori mai lung decât cel inițial, cu parametri statistic imperceptibili ai repartițiilor factorilor examinați; s-a ilustrat că aplicarea acestei metode permite aprecierea mai exactă a mărimii legăturii liniare corelaționale, în comparație cu procedurile analizei *bootstrapping*; analizată calitatea modelelor, construite cu utilizarea procedurii de mărire virtuală a dimensiunii eșantionului, dedusă concluzia că aplicarea procedurii măririi virtuale a dimensiunii eșantionului, bazate pe metoda distribuției punctuale, într-un șir de cazuri ajută la obținerea unor rezultate mai calitative de modelare matematică, ceea ce permite cercetătorului să ia decizii mai eficiente în analiza datelor multidimensionale de volum mic.

Secțiunea șase conține concluziile deduse în baza rezultatelor capitolului, în care se spune, că în conformitate cu datele obținute este necesar de formulat metodica de construire SSD după eșantioane pasive de volum mic pentru cazurile când metodele clasice nu permit obținerea modelului de calitate necesară în baza datelor inițiale. În legătură cu aspectul laborios în procesarea eșantioanelor de volum mic (aplicarea metodei de distribuție punctuală) și cu complexitatea compunerii tablei virtuale de date pentru modelarea ulterioară, este necesar de a elabora un complex de produse program specializat, care realizează metodele propuse. Se recomandă de verificat metodicile elaborate și de elaborat SSD pentru diferite domenii de activitate a cercetătorilor.

În *al treilea capitol* «Sisteme suport de luare a deciziilor în baza eșantioanelor de volum mic» se formează, realizează și se verifică metodica de construire a SSD după eșantioanele pasive de volum mic.

În *prima secțiune*, în baza informației analizate și datelor obținute, se formulează metoda, care permite asigurarea aparatului matematic de suport în luarea deciziilor la procesarea tabelelor experimentului pasiv, în cazul cantității mici de date inițiale.

Vom prezenta integral acest algoritm:

1. Depistarea datelor aberante. După cum este arătat în secțiunea 2.1, în calitate de procesare preventivă a tabelelor experimentului pasiv, la etapa analizei datelor obținute, în cazul prezenței informației prealabile despre tipul legii de distribuție a valorilor aleatorii, care intră în spațiul factorial, la volumul eșantionului $n \leq 10$, se recomandă aplicarea criteriului Dixon pentru identificarea măsurărilor anormale, ce pot fi legate de oboseala personalului, dereglările utilajului, erori în procesul tehnologic și altele. În cazul lipsei informației prealabile necesare și imposibilitatea de a determina cu precizia necesară tipul legii de distribuție a valorii aleatorii, de asemenea este posibil de a identifica valorile aberante prin criteriul Lvovski. La volumul eșantionului $n \geq 10$ căutarea rezultatelor anormale este posibil de efectuat prin unul din următoarele criterii:

1. Metoda Irwin.
2. Criteriul Lvovski.
3. Criteriul Tietjen-Moore.
4. Criteriul Dixon (pentru distribuție normală).

În afară de aceasta, cum este arătat în partea teoretică a cercetării, aplicarea criteriului Tietjen-Moore dispune de o cantitate mai mică de erori de tipul I, de aceea utilizarea lui se recomandă pentru confirmarea ipotezei despre prezența valorii aberante în datele de volum mic.

În cazul detectării datelor aberante în una din valorile cercetate, din tabelul inițial al datelor pasive multidimensionale se elimină tot rândul care conține măsurarea anormală. Dacă eliminarea rândului nu este posibilă din cauza cantității reduse a informației inițiale, valoarea poate fi înlocuită cu una apropiată după caracteristicile statistice ale grupului principal de date.

2. Procesarea ulterioară a tabelului are drept scop identificarea perechilor de date aberante bidimensionale. Cu acest scop este necesar de construit tabelele distribuției bidimensionale pentru fiecare din perechile valorilor de intrare X_j . Analizând tabelele din punct de vedere al prezenței valorilor extreme, în cazul identificării lor, de asemenea este necesar de a șterge din eșantionul inițial rândurile în care ele se conțin. Este necesar de remarcat că stabilirea prezenței datelor aberante pereche este necesar de efectuat în cazul dacă datele inițiale au fost obținute în timpul unui oarecare proces tehnologic, unde condițiile obținerii datelor fac posibile erorile de acest tip.

3. Următoarea etapă a algoritmului ține de elaborarea variantelor de rezolvare a problemei examinate. Pregătirea către analiza de regresie ulterioară trebuie să înceapă de la găsirea valorii legăturii liniare corelaționale între factorii care intră în tabelul experimentului pasiv. Pentru aceste scopuri este necesar de efectuat analiza corelațională prin metoda pleiadelor corelaționale, cu coeficientul de prag al corelării Pearson $r=0,75$. Pentru determinarea factorilor, care caracterizează pleiadele, pot fi utilizate metode expert de evaluare. La imposibilitatea determinării factorilor de caracterizare, sau când au valori egale pentru cercetare, este recomandat de a determina toate variantele posibile de combinare a factorilor slab legați pentru analiza de regresie ulterioară și de a forma din ele diverse tabele pentru procesarea lor ulterioară.

4. După selectarea din tabel a datelor factoriale inițiale, legate una cu alta prin dependență liniară slabă, se efectuează procedura măririi volumului datelor multidimensionale, bazată pe aplicarea metodei de distribuție punctuală pentru fiecare din factorii variabili de intrare și de ieșire, precum și aprecierea valorilor densităților probabilităților nenormate în domeniul virtual. În aceste scopuri este utilizată informația prealabilă despre tipul distribuției valorii aleatorii, cunoscută din caracteristicile procesului studiat, apreciată estimativ după histogramă, sau prealabil calculată prin una din metodele de obținere a aprecierilor pe eșantioane mici.

Drept rezultat al acestei operații servește formarea tabelului de date mărite virtual, care reflectă informația din eșantionul multidimensional prealabil, care este mai lung decât cel inițial de 10-20 ori, în dependență de calitatea datelor prelabile și legea distribuției valorilor aleatorii ale factorilor.

5. Conform tabelului datelor mărite virtual în mod repetat se identifică coeficienții de corelare pentru toți factorii, după principiul «fiecare cu fiecare», după care se efectuează analiza rezultatelor obținute prin metoda pleiadelor corelaționale la valoarea de prag a coeficientului Pearson $r=0,8$. În cazul identificării suplimentare a factorilor legați prin dependență liniară puternică, de asemenea este recomandat să fie excluși din examinarea ulterioară, lăsând doar câte un factor, care cel mai exact caracterizează fiecare din pleiade.

6. După evidențierea din eșantionul mărit virtual a factorilor slab legați, stabilim variantele posibile a modelelor matematice, care conțin în sine eșantionul virtual multidimensional.

7. Pentru acest experiment construim modelele matematice în baza tabelului virtual de date, obținut prin oricare din metodele analizei de regresie (cunoscute și convenabile) ale datelor pasive multidimensionale.

8. Verificăm adecvarea și calitatea modelelor construite, determinăm modelele matematice optimale, trecem la interpretarea lor, împreună cu specialiștii de profil și domeniul concret de cercetare.

9. Generalizăm informația obținută în SSD, bazată pe datele colectate, modelul matematic al procesului studiat și cunoștințele formale ale experților în domeniul cercetat.

În *secțiunea a doua* se examinează realizarea metodicii obținute în formă de produse program; se elaborează cerințele față de produsele program; se efectuează alegerea metodelor și mijloacelor pentru realizarea lor; se descriu componentele ce fac parte din complexul de program, care realizează metoda distribuției punctuale, metoda pleiadelor corelaționale și metoda de îmbinare a tabelor densității nenormate a probabilității în domeniul virtual. Produsul program pentru modelarea matematică și analiza datelor pe eșantioane pasive de volum mic constă din trei programe independente una față de alta, unite prin interfețe comune de lucru cu datele de intrare și ieșire [19, 20]. Fiecare subprogram este o aplicație consolă aparte, care lucrează cu fișiere *csv*, cea ce simplifică implementarea lor în diverse aplicații.

În *secțiunea trei* este ilustrată aplicarea metodicii obținute pentru prognozarea complexității operațiilor laparoscopice la bărbați cu colecistite acute, ceea ce în baza analizei datelor inițiale ale examinării pacientului lărgeste posibilitățile chirurgului în prognozarea dificultăților intraoperatoriale, care ar putea să intervină. În acest compartiment este efectuată clasificarea complexității operațiilor, propusă de experți. Pe parcursul pregătirii analizei inițiale a datelor au fost efectuate verificările prezenței datelor aberante în datele inițiale, precum și normalitatea valorilor cercetate. Este examinată problema determinării mărimii coeficientului liniar corelațional în eșantioane de volum mic, stabilite avantajele determinării mărimii coeficientului de corelare Pearson după mărirea virtuală a eșantionului prin metoda multidimensională de distribuție punctuală. În baza modelului obținut este elaborat sistemul computerizat de suport în luarea deciziilor. Sistemul a fost aprobat pe operațiile efectuate pe parcursul următorului an după implementare: modelul a identificat corect complexitatea operației ce urma să fie efectuată în 73 % din cazuri, ceea ce poate fi considerat ca satisfăcător pentru prognozarea timpurie. Verificarea modelului obținut pe setul final de date a arătat că modelul este unul adecvat, statistic semnificativ, iar factorii ce îi conține descriu corect valoarea de ieșire aproximativ în 71%, la mărirea erorii medii pătratice egale cu 0,712. Este efectuată comparația valorilor coeficientului de corelare mixtă (r^2), erorii medii pătratice de apreciere (S_{xy}) și a valorii (F) a criteriului Fischer, calculate în baza setului final de date, pentru modelul obținut și pentru modelele care ar putea fi construite fără aplicarea metodei propuse. Se constată că modelul obținut dispune de cele mai bune caracteristici, ceea ce înseamnă că este cel mai calitativ. Deoarece sistemul utilizează metode probabilistice și statistice, trebuie să

remarcăm că nivelul prezis de complexitate a intervenției chirurgicale întotdeauna are o anumită parte de eroare, în legătură cu aceasta, de rând cu cel mai probabil rezultat, pentru decident se afișează și recomandările pentru o finalizare a evenimentelor mai puțin probabilă. În așa fel, în fereastra cu recomandări în partea de sus se indică nivelul de complexitate al operației, calculat în baza datelor introduse, după ce valoarea se rotunjește și în blocuri speciale de text se afișează informația despre descrierea nivelului prezis de complexitate a operației, precum și recomandările de realizare a ei. În partea de jos a ferestrei se indică descrierea și recomandările pentru efectuarea operației în cazul unei situații mai puțin probabile, pentru ca decidentul în baza cunoștințelor și experienței sale să poată lua cea mai potrivită decizie.

În *secțiunea patru* se examinează aplicarea metodicii obținute la crearea sistemului computerizat orientat spre problemă, asigurând un suport la luarea deciziilor despre calitatea coacerii de probă a produselor de panificație. Aprecierea calității producției finale și emiterea recomandărilor de ajustare operativă a rețetei și procesului tehnologic în baza partidei de probă este o procedură analitică dificilă. În același timp, deoarece se realizează nu mai mult de 5 cicluri de producere pentru fiecare tip de producție, iar parametrii tuturor produselor într-o coacere de probă sunt identici, ceea ce este caracteristic pentru orice proces de grup, rezultă că datele în baza cărora trebuie luată decizia sunt și un eșantion mic, și un plan suprasaturat. În secțiunea aceasta se descriu principalii parametri de intrare și ieșire, care influențează calitatea produselor de panificație, ce se referă atât la rețetă cât și la procesul tehnologic. În baza exemplului produsului numit «Pâine curativă» este ilustrat procesul de analiză a datelor, de construire a modelelor matematice și emitere a recomandărilor de către sistemul elaborat. Se stabilește că aplicarea măririi virtuale a datelor, bazată pe metoda distribuției punctuale, poate îmbunătăți calitatea analizei statistice a tabelelor obținute la producerea produselor de panificație. În afară de cunoștințele despre parametrii cantitativi, care se obțin cu ajutorul modelării, în dicționarul termenilor SSD proiectate de asemenea trebuie de adăugat cunoștințele obținute din documentele normativ-tehnice, precum și de la experții din acest domeniu. În baza acestor cunoștințe a fost formulat un set de reguli, care leagă caracteristicile separate ale procesului tehnologic și rețetele, permițând explicarea anumitor fenomene, precum și prognozarea comportării parametrilor studiați. Marea parte a regulilor incluse în baza de cunoștințe sunt cele de producție. Astfel, partea stângă a regulii prezintă în sine un set de premise, iar în partea dreaptă se conține concluzia, care poate fi făcută în baza lor. O asemenea regulă lucrează, dacă sunt îndeplinite premisele legate între ele prin setul de operații booleene. În același timp fiecare premiză este o valoare determinată, exprimată cantitativ, valoarea unuia din parametrii cercetați. În așa fel, în SSD elaborat predicția tuturor parametrilor cantitativi se bazează pe modelul matematic, construit în conformitate cu metodele și algoritmi elaborați. Însă

o parte din parametri se analizează doar pe cale organoleptică și sunt apreciați prin punctaj, ceea ce nu permite aplicarea aparatului statistico-matematic asupra lor. Pentru gestionarea varietăților acestor date se aplică reguli de producție, care analizează atât indicii cantitativi cât și indicii apreciați prin punctaj [17, 30]. În secțiune de asemenea se descrie logica funcționării și interfața produsului program.

Secțiunea a cincea este dedicată concluziilor și propunerilor conform rezultatelor capitolului, în care se subliniază următoarele: aplicarea metodicii elaborate pentru eșantioanele de volum mic, obținute în diferite condiții, în multe cazuri permite îmbunătățirea relativă a calității modelelor matematice obținute, păstrând legea primară de distribuție a valorilor aleatorii; aplicarea metodicii elaborate, fără a dispune de informația prealabilă despre tipul legii de distribuție, în baza analizei histogramelor, sau emiterea presupunerii despre tipul legii de distribuție, de asemenea poate ajuta la îmbunătățirea calității modelelor matematice obținute; aplicarea măririi virtuale a informației ținând cont de tipul legii de distribuție, în baza metodei de distribuție punctuale, într-un șir de cazuri permite identificarea mai precisă a valorii legăturii liniare corelaționale prin coeficientul de corelare Pearson; pentru obținerea modelelor matematice statistic semnificative este necesar de aplicat diferite metodici de construire a ecuațiilor de regresie; produsele program elaborate permit accelerarea procesării datelor pasive de volum mic conform metodicii elaborate, executând cele mai grele părți de calcul și transformare a datelor; produsul program elaborat și utilizarea sa pentru suportul deciziilor în domenii aplicative concrete de cercetare confirmă rezultatele obținute în teză.

CONCLUZII ȘI RECOMANDĂRI GENERALE

În rezultatul cercetării efectuate:

1. Au fost analizate criteriile existente pentru identificarea datelor aberante în eșantioane de volum mic [12], au fost identificate metodele care dispun de cea mai mare putere [16], au fost expuse recomandări pentru identificarea datelor aberante în eșantioane de volum mic [18].

2. A fost elaborată metodica de mărire virtuală a tabelului datelor pasive multidimensionale în baza metodei de distribuție punctuale [29], folosind informația despre tipul legii distribuției fiecărei valori aleatoare unidimensionale [24] și valoarea densității nenormate a probabilității în domeniul virtual. Metodica elaborată permite obținerea tabelului de date, care este statistic omogen cu cel primar, păstrând tipul legii distribuției fiecărei valori unidimensionale separate aleatoare, iar numărul de rânduri fiind mai mare decât cel inițial de 10-20 ori [31].

3. Au fost analizate metodicile existente pentru determinarea valorii legăturii liniare corelaționale în eșantioane de volum mic, s-a stabilit că niciuna din ele nu dispune de o precizie necesară pentru efectuarea analizei prin metode clasice, în legătură cu aceasta fiind recomandată utilizarea valorii de prag pentru identificarea dependenței liniare puternice prin coeficientul Pearson, care este egal cu $r=0,75$ [14].

4. S-a arătat că aplicarea măririi virtuale a tabelului de date pasive permite determinarea mai precisă a valorii legăturii liniare de corelare [13].

5. S-a elaborat algoritmul construirii sistemelor suport pentru decizii, bazat pe metodele probabilistice și statistice elaborate, care permite îmbunătățirea calității modelelor de regresie prin eșantioane de volum mic, în cazurile când metodele clasice nu oferă exactitatea necesară [7, 15].

6. S-a elaborat produsul program care permite simplificarea și accelerarea considerabilă a procesului de procesare a eșantioanelor multidimensionale de volum mic în baza metodei de distribuție punctuală [19, 20].

7. Metodica a fost testată în diferite domenii, pe eșantioane obținute în diferite condiții. Au fost construite două SSD orientate pe probleme, pentru domenii în care este imposibilă obținerea unor volume mari de date experimentale. Rezultatele obținute au arătat că aplicarea metodelor elaborate poate permite construirea sistemelor, care nu numai vor răspunde cerințelor înaintate pentru calitatea deciziilor, dar și într-un șir de cazuri vor conduce spre obținerea unor rezultate mai justificate în comparație cu aplicarea metodelor clasice [17, 21, 22].

8. Sistemele suport de luare a deciziilor, elaborate de autor, au permis:

8.1. Emiterea recomandărilor despre operațiile laparoscopice planificate, nivelul lor de dificultate, ordinea de efectuare și probabilitatea complicațiilor, recomandările fiind adresate intervalului de timp, când nu a fost posibilă colectarea unui volum suficient de date.

8.2. Aprecierea calității coacerii de probă a produselor de panificație pe baza generării interactive a modelelor și elaborarea variantelor de decizie adaptate la situația concretă.

SSD elaborate pot fi cu succes integrate în sisteme informaționale mai ample, prin utilizarea și modernizarea bazelor de date create și a modelelor matematice propuse.

Prin urmare, problema cercetată în cadrul acestui studiu, constă în determinarea metodicilor probabilistice și statistice de suport la luarea deciziilor în baza construirii modelelor matematice adecvate prin eșantioane pasive multidimensionale de volum mic, obținute în diferite condiții.

Realizarea studiului asupra suportului luării deciziilor pe eșantioane pasive de volum mic a permis formularea recomandărilor de bază, care pot fi utile atât din punct de vedere științific cât și practic:

1. Luarea deciziei despre prezența datelor aberante în datele de dimensiuni mici trebuie efectuată cu ajutorul unuia sau a câtorva criterii: Irwin, Lvovski, Tietjen-Moore sau Dixon [16, 18]. Pe lângă aceasta trebuie aditional de studiat influența devierii de la legea normală a distribuției asupra puterilor acestor criterii, precum și posibilitatea identificării erorilor grupate și variantelor de eliminare a lor.

2. Pentru îmbunătățirea calității deciziilor aplicate la datele de dimensiuni mici se recomandă utilizarea metodicii propuse de mărire virtuală a dimensiunii eșantionului [29], care permite păstrarea cunoștințelor despre tipul legii distribuției pentru fiecare din valorile aleatoare studiate.

3. După mărire virtuală a dimensiunii eșantionului, bazată pe aplicarea metodei distribuției punctuale, se recomandă efectuarea analizei corelaționale cu coeficientul de prag Pearson egal cu 0,8 [13]. Trebuie de remarcat că nu au fost identificate metode, care ar permite efectuarea analizei corelaționale de date cu dimensiune mică cu precizia și veridicitatea necesară [14], în legătură cu aceasta o direcție a cercetărilor viitoare poate deveni elaborarea metodicii pentru estimarea valorii legăturii între eșantioanele mici.

4. Se recomandă aplicarea rezultatelor obținute în scopul îmbunătățirii calității elaborării sistemelor suport pentru decizii pentru probleme concrete în baza analizei probabilistice și statistice a volumurilor mici de informație, indiferent de domeniul de cercetare, dacă acumularea cantității necesare de date inițiale nu este posibilă sau nu este economic eficientă.

БИБЛИОГРАФИЯ

1. Гаскаров Д. В., Шаповалов В. И. Малая выборка. Москва: Статистика, 1978, 248с.
2. Григорьев Ю.Д. Метод Монте-Карло: вопросы точности асимптотических решений и качества генераторов псевдослучайных чисел. В: Заводская лаборатория. Диагностика материалов, 2016, т.82, №7, с. 72-84.
3. Громыко Г.Л. Теория статистики: Учебник, 2-е изд., перераб. и доп. Москва: ИНФРА-М, 2011, 476 с.
4. Долгов Ю.А. Статистическое моделирование: Учебник для вузов. – 2-е изд., доп, Тирасполь: Полиграфист, 2011, 352 с.
5. Еременко В.С., Куц Ю.В., Мокийчук В.М. Оценка однородности выборок малого объема, В: Системи обробки інформації, 2006, №7(56), с.26-29.
6. Кобзарь А.И.. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2012, 816 с.
7. Кожокару С.К., Попукайло В.С. Метод принятия решений в условиях малой выборки. В: Научный обозреватель, 2016, №3(63), с.79-80.
8. Козыр Е.В. Этапы обработки данных при малых выборках. В: Механіка гіроскопічних систем, 2010, №21, с.103-111.
9. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло. В: Научный журнал КубГАУ, 2015, №114(10), с.27-41.
10. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. Ростов-на-Дону: Феникс, 2009, 572с.
11. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? В: Заводская лаборатория. Диагностика материалов, 1991, т.57, №7, с.64-66.
12. Попукайло В.С. Исследование критериев грубых ошибок применительно к выборкам малого объема В: Радіоелектронні і комп'ютерні системи, 2015, №3(73), с. 39-44.
13. Попукайло В.С. Исследование линейной корреляционной связи в многомерном методе точечных распределений В: Информационно-управляющие системы, 2016, №6, с. 96-98. doi:10.15217/issn1684-8853.2016.6.96
14. Попукайло В.С. Исследование линейной корреляционной связи в парных выборках малого объема. В: Технология и конструирование в электронной аппаратуре, 2016, №1, с. 27-32. doi: 10.15222/ТКЕА2016.1.27

15. Попукайло В.С. Методы математического моделирования по выборкам малого объема В: Доклады итоговой научн. конф. проф.-препод. состава инженерно-технического институт за 2015 год: г. Тирасполь 22-26 янв. 2016г./ Приднестр. гос. ун-т им. Т.Г. Шевченко; под ред.: Долгова Ю.А. Тирасполь: «TesLine», 2016, с. 150-151.

16. Попукайло В.С. Обнаружение аномальных измерений при обработке данных малого объема. В: Технология и конструирование в электронной аппаратуре, 2016, №1, с. 27-32. doi: 10.15222/ТКЕА2016.4-5.42

17. Попукайло В.С. Поддержка принятия решения о качестве пробной лабораторной выпечки. В: Проблемы современной науки и образования, 2017, №7(89) с. 43-45 doi:10.20861/2304-2338-2017-89-003

18. Попукайло В.С. Принятие решений о наличии выбросов в данных малого объема. In: Materialele Conferinței Științifice a Doctoranzilor (cu participare internațională) “Tendințe contemporane ale dezvoltării științei: viziuni ale tinerilor cercetători”, 25 mai 2016, Chișinău, Universitatea Academiei de Științe a Moldovei, 2016, p.306-310.

19. Попукайло В.С. Построение таблицы исходных данных для многомерного моделирования по выборкам малого объема. В: Тезисы IX Международной конференции «Математическое моделирование в образовании, науке и производстве». Тирасполь, 8-10 октября 2015г. Тирасполь: Изд-во Приднестровского университета, 2015, с.176-177.

20. Попукайло В.С. Программная реализация метода точечных распределений. В: Сборник материалов Всерос., научно-практической конференции «Информационно-телекоммуникационные системы и технологии», 16-17 окт. 2015 г., Кемерово [Электронный ресурс] / ФГБОУ ВПО «Кузбас. гос. техн. ун-т им. Т. Ф. Горбачева»; редкол.: Трофимов И.Е. (отв. ред.) [и др.], Кемерово, 2015. <http://sibscience.ru/page/ITSIT-2015/ITSIT/6-Matematicheskoe-modelirovanie/6042.pdf>

21. Попукайло В.С. Построение математической модели эффективности работы банка в условиях малой выборки. В: Актуальные проблемы гуманитарных и естественных наук, 2016, №3(86), с. 74-78.

22. Попукайло В.С. Прогнозирование уровня успеваемости абитуриентов в условиях малой выборки. В: Наука вчера, сегодня, завтра. Сб. статей по материалам XXXII международно-практической конференции, №3(25), 2016, с. 95-101.

23. Столяренко, Ю.А. Контроль кристаллов интегральных схем на основе статистического моделирования методом точечных распределений. Диссертация на соискание ученой степени к.т.н. М., ГУП НПЦ «СПУРТ», 2006. 191 с.

24. Столяренко Ю.А. Гординский Д.Н. Попукайло В.С. Применение метода точечных распределений для различных законов распределения В: Тезисы VII Международной конференции. 3-5 октября 2013г. – Тирасполь: Изд-во Приднестровского университета, 2013, с. 201-202.
25. Gaındric C. Luarea decizilor. Metode si tehnologii. Chisinau: Stiinta, 1998, 162p.
26. Fischer H. A History of the Central Limit Theorem: From Classical to Modern Probability Theory, Sources and Studies in the History of Mathematics and Physical Sciences, New York: Springer, 2011, p.402.
27. Gurov S. 0-event: Point and Interval Estimates. In: Mathematics and Statistics, 2013, №1(2): p. 53-58
28. Hays W. Statistics. Boston: Cengage Learning, 1994, 1136 p.
29. Popukaylo V. Analysis of statistical modeling methods for small-volume samples. In: Proceedings Proceeding of the 5th International Conference «Telecommunications, Electronics and Informatics» ,may 20-23, 2015, Chişinău, Tehnica-UTM, Moldova, 2015, p.207-211.
30. Popukaylo V. Development of the decision support system for test baking. In: Proceedings CMSM 4, June 28 – July 2, 2017 Chişinău / ed.: Mitrofan Choban [et al.]. Chişinău: Institute of Mathematics and Computer Science, 2017 (CEP USM), 572 p.
31. Popukaylo V. Small size sample mathematical modeling B: Meridian Ingineresc 2015, №4(595), с. 25-30.

АННОТАЦИЯ

Попукайло Владимир

ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ ПО ПАССИВНЫМ ВЫБОРКАМ МАЛОГО ОБЪЕМА

диссертация доктора информатики, Кишинев, 2017.

Структура диссертации: введение, три главы, общие выводы и рекомендации, библиографический список из 160 наименований, 4 приложения, 120 страниц базового текста, 30 рисунков, 23 таблицы. По материалам диссертационного исследования опубликовано 16 печатных работ.

Ключевые слова: теория принятия решений, системы поддержки принятия решений, малая выборка, пассивный эксперимент, корреляционный анализ, регрессионный анализ, грубые промахи, метод точечных распределений.

Область исследования: пассивные малые выборки, полученные в различных условиях.

Цель исследования: разработка методик построения систем поддержки принятия решений по пассивным выборкам малого объема, в условиях, когда классические вероятностно-статистические методы не позволяют получить обоснованных выводов.

Задачи исследования: 1) проанализировать различные подходы и методы обработки выборок малого объема; 2) исследовать возможности использования существующих критериев для нахождения грубых промахов в выборках малого объема; 3) исследовать возможности критериев обнаружения линейной корреляционной связи в выборках малого объема; 4) разработать методику построения систем поддержки принятия решений в условиях пассивного эксперимента и невозможности получить большой массив исходных данных; 5) апробировать разработанную методику на данных, полученных в различных условиях.

Научная новизна и оригинальность исследования – разработана оригинальная методика, позволяющая строить системы поддержки принятия решений по пассивным выборкам малого объема, дающие надёжные рекомендации для лица, принимающего решения.

Актуальность исследования состоит в том, что существует большое количество областей, в которых невозможно получить данные большого объема, в связи с чем, принятие решения с использованием классических методик крайне затруднено или невозможно.

Важная научная проблема, решенная в работе: определение вероятностно-статистических методик поддержки принятия решений, на основании выборок малого объема, полученных в ходе пассивного эксперимента.

Теоретическая значимость исследования заключается в методике улучшения качества принимаемых решений на основе предложенных подходов и алгоритмов, опирающихся на вероятностно-статистические методы обработки информации.

Практическая значимость результатов работы. Предложенные в диссертационной работе методики поддержки принятия решения могут быть широко применены при статистическом анализе в различных областях исследований, в которых нет возможности получить большое количество данных.

Внедрение результатов исследования: Результаты работы внедрены в лечебный процесс отделения Эндоскопической и малоинвазивной хирургии Республиканской клинической больницы (г. Тирасполь), учебный процесс кафедр «Биологии и физиологии человека» и «Терапии №2» медицинского факультета Университета им. Т.Г. Шевченко (г. Тирасполь), в работу научно-исследовательской лаборатории «Математическое моделирование» и в учебную программу магистратуры по направлению «Информатика и вычислительная техника» того же университета.

ADNOTARE
Popukaylo Vladimir

SUPPORTUL DECIZIILOR ÎN BAZA EȘANTIOANELOR PASIVE DE DIMENSIUNI MICI

Teză de doctor în informatică. Chișinău, 2017

Structura tezei: introducere, trei capitole, concluzii, bibliografie – 160 titluri, 4 anexe, 120 pagini de text de bază, 30 figuri, 23 tabele. Rezultatele cercetării sunt publicate în 16 lucrări științifice. **Cuvinte-cheie:** teoria luării deciziilor, sisteme pentru suport decizii, eșantioane mici, experiment pasiv, analiză corelațională, analiză regresivă, date aberante, metoda distribuțiilor punctate.

Domeniul de cercetare: eșantioane mici pasive, obținute în diferite condiții.

Scopul lucrării: elaborarea metodicii de construire a sistemelor pentru suport decizii în baza eșantioanelor de dimensiuni mici, în condițiile , când metodele clasice ale teoriei probabilităților și statisticii matematice nu permit obținerea unor concluzii fundamentate.

Obiectivele cercetării: 1) analiza diferitelor abordări și metode de procesare a eșantioanelor de dimensiuni mici; 2) cercetarea posibilității de utilizare a criteriilor existente pentru identificarea datelor aberante în eșantioane de volum mic; 3) cercetarea posibilităților de determinare a relației de corelare liniară în eșantioane de volum mic; 4) elaborarea metodicii de construire a sistemelor pentru suport decizii în condiții de experiment pasiv și imposibilitatea obținerii unui masiv mare de date inițiale; 5) aprobarea metodicii elaborate pentru date, obținute în diferite condiții; 6) fundamentarea și demonstrarea aplicabilității metodologiei propuse.

Noutatea științifică a lucrării o constituie metodologia originală de construire a sistemelor pentru suport decizii în baza eșantioanelor pasive de dimensiuni mici, care-i oferă decidentului cele mai exacte recomandări.

Actualitatea cercetării este motivată de existența unui spectru larg de domenii pentru care este imposibil de obținut un volum mare de date, fapt ce împiedică sau chiar face imposibilă luarea deciziilor prin aplicarea metodelor clasice.

Problema științifică principală, soluționată de către autor, constă în elaborarea metodicii probabilistice și statistice de suport al deciziilor în baza eșantioanelor de dimensiuni mici, obținute în baza experimentului pasiv.

Semnificația teoretică a cercetării o constituie metoda de îmbunătățire a calității deciziilor prin abordările propuse și algoritmilor elaborați, bazați pe metode probabilistice și statistice de procesare a informației.

Valoarea aplicativă a cercetării: metodele propuse în teză pot fi utilizate pe larg în analiza statistică în diverse domenii, pentru care nu există posibilitatea de a obține un volum mare de date.

Implementarea rezultatelor cercetării: rezultatele cercetării sunt implementate în procesul de tratament al Secției de chirurgie endoscopică și minim invazivă al spitalului clinic republican din or. Tiraspol, procesul didactic al catedrelor “Biologia și fiziologia omului” și “Terapia №2” a facultății de medicină a Universității “T.G.Șevcenco” din or. Tiraspol, precum și în cercetările din cadrul laboratorului de modelare matematică și cursurile de masterat la specialitatea „Informatică și tehnică de calcul” ale aceleiași universități.

ANNOTATION

Popukaylo Vladimir

DECISION SUPPORT ON SMALL SIZE PASSIVE SAMPLES

Thesis of Doctorin Computer Science. Kishinev, 2017.

Structure of the thesis: Introduction, three Chapters, General conclusions and recommendations, bibliography of 160 titles, 4 appendices, 120 pages of body text, 30 figures, 23 tables. According to the dissertation research materials 16 printing works are published.

Keywords: mathematical modeling, small sample, passive experiment, correlation analysis, regression analysis, outliers, the method of point distributions.

Field of study: small size passive samples obtained in different conditions.

Research objective: to develop a construction technique of adequate mathematical models for small size passive samples, in conditions when classical probabilistic-statistical methods do not allow obtaining valid conclusions.

Research tasks: 1) to analyze various approaches and methods of small size samples processing; 2) to investigate the possibility of using existing criteria for outliers finding in small size samples; 3) to investigate the possibility of the linear correlation detection in the small size samples; 4) to develop a methodology for constructing decision support systems in cases of passive experiment and impossibility of obtaining a large amount of raw data; 5) to test the developed technique on the obtained data in various conditions; 6) to substantiate and to prove the applicability of the developed technique.

The scientific novelty and originality of the research: the original methodology for construction the decision support systems which give the most accurate recommendations for decision-makers basing on passive small samples.

The relevance of research: there are many areas where it is impossible to obtain large amounts of data, and therefore, the decision-making using classical techniques is extremely difficult or impossible.

Important scientific problem solved in the work: the determination of probabilistic and statistical methods of decision support based on small size samples obtained during the passive experiment.

The theoretical significance of the study lies in the methodology of improving the quality of decisions taken on the basis of the proposed approaches and algorithms based on probabilistic and statistical methods of data processing

The practical significance of the research results. The decision support methods, proposed in the thesis, can be widely used in the statistical analysis in various fields of research, where it is impossible to obtain a large amount of data.

The research results implementation: The results of the research is implemented in the direct care of Department of Endoscopic and Minimally Invasive Surgery of the Republican Clinical Hospital (Tiraspol), into educational process of the department "Biology and Human Physiology" and "Therapy №2» Medical Faculty of the T.G. Shevchenko University, into the work of the Research Laboratory "Mathematical Modeling" and academic Master's program in "Computer Science and Engineering" of the same university.

POPUKAYLO VLADIMIR

**SUPPORTUL DECIZIILOR ÎN BAZA EȘANTIOANELOR PASIVE
DE DIMENSIUNI MICI**

122.03 – MODELARE, METODE MATEMATICE, PRODUSE PROGRAM

Autoreferatul tezei de doctor în informatică

Aprobat spre tipar: 05.06.2017

Formatul hârtiei 60x84 1/16

Hârtie ofset. Tipar ofset.

Tiraj 50 ex.

Coli de tipar: 1,75

Comanda nr. ____

Tipografia «ВлаСаНа» SRL,
MD-3200, or. Bender, str.Lenin, 29.