

## STATISTICA ȘI LOCUL EI ÎN ECONOMIA CUNOAȘTERII CA INSTRUMENT DE GÂNDIRE

<sup>1</sup>*Dr., prof. univ., Ion PÂRȚACHI*

<sup>2</sup>*Dr., prof. univ., Alexei LEAHU*

<sup>1,2</sup>*Academia de Studii Economice a Moldovei,  
Republica Moldova, Chișinău, Bănulescu Bodoni, 61,  
tel. (+373) 22 41 28, [www.ase.md](http://www.ase.md)*

### **Abstract**

The paper focuses on several examples of concrete applications of statistics that confirm its quality as a tool of intelligent thinking in the knowledge economy. Furthermore, two of the main principles on which this thinking is based are emphasized: the principle of statistical regularity and the principle of Maximum Likelihood.

**Key words:** *dispersion analysis, mean value, hypergeometric distribution, Maximum Likelihood, principle of statistical regularity.*

**JEL CLASSIFICATION:** C1, C13

Tema lucrării noastre este inspirată de Richard E. Nisbett prin cartea sa *Mindware: Instrumente pentru o gândire inteligentă*, București, Ed. Litera, 2016. Autorul ei, profesor de psihologie la University of Michigan (SUA), este unul din cei mai respectați psihologi la nivel mondial. Dat fiind faptul ca Nisbett aduce in calitate de exemplu de instrument de gândire inteligentă (pe langa multe altele) si *Statistica*, noi vom incerca sa ne oprim mai amanuntit anume asupra acesteia, deoarece gândirea inteligentă ține, fără îndoială, de economia cunoașterii.

Trăim în mileniul în care previziunea ce vizează Statistica, previziune făcută de scriitorului englez *Herbert G. Wells (1866-1946)*, devenit celebru pentru cărțile sale de ficțiune, s-a adeverit. Acesta spunea, citez: "Modul de gândire statistic va deveni, într-o zi, la fel de necesar precum abilitatea de a citi ori de a scrie". Nu e de mirare, de exemplu, că în Franța, *Statistica Descriptivă*, iar în SUA, *Elementele de Teorie a Probabilităților [2]* au fost introduse în programul de studiu gimnazial cu mai mult de jumătate de secol în urmă. Or, în mileniul III, problemele cu care se confrunta managerii, oamenii de afaceri in general, nu mai sunt cele referitoare la disponibilitatea ori la lipsa informațiilor ci, mai degrabă, cele legate de modul in care să folosească informațiile in scopul de a lua decizii mai bune. Orice specialist angrenat în economia cunoașterii, care stăpânește și utilizează Statistica în activitatea sa, ne poate confirma că acest instrument de lucru ne ajută în:

- a descrie și a prezenta in modul cel mai compact și potrivit datele/informațiile cu care lucrăm;
- a analiza și interpreta aceste date/informații in vederea validării/invalidării unor ipoteze sau modele matematice de lucru;
- a obține previziuni credibile privind variabilele de interes;
- a îmbunătăți desfășurarea activităților de care sunteți răspunzători;
- a trage concluzii privind colectivități numeroase, având la dispoziție date obținute din esantioane. Cu alte cuvinte, studiind partea, putem cunoaște întregul.

În aceeași ordine de idei se inscrie ca o mare provocare pentru știința Statistică și problema cunoscută sub denumirea BIG DATA. Este vorba despre datele păstrate și prelucrate în cantități imense, datorită unor medii de stocare mai ieftine, unor metode de procesare mai rapide și unor algoritmi mai performanți. BIG DATA a devenit o problemă în afaceri, sau cel puțin o problemă pe care oamenii de afaceri incep să o conștientizeze. Presa începe să aloce din ce în ce mai mult spațiu acestui subiect. Plecând de la Wall Street Journal "Companiile sunt inundate cu date" ("Companies are being inundated with data"), continuând cu Financial Times "Din ce în ce mai mult în afaceri

sunt aplicate analize din mass-media, ca Facebook și Twitter" ("Increasingly businesses are applying analytics to social media such as Facebook and Twitter"), și terminând cu Forbes "Big Data a ajuns la Seton Health Care Family" ("Big Data has arrived at Seton Health Care Family"). De ce atâtea articole pe aceasta temă? Deoarece BIG DATA are potențialul de a afecta profund modul în care facem afaceri și chiar modul nostru de a trăi, dar și va avea un impact profund în Economia Cunoașterii.

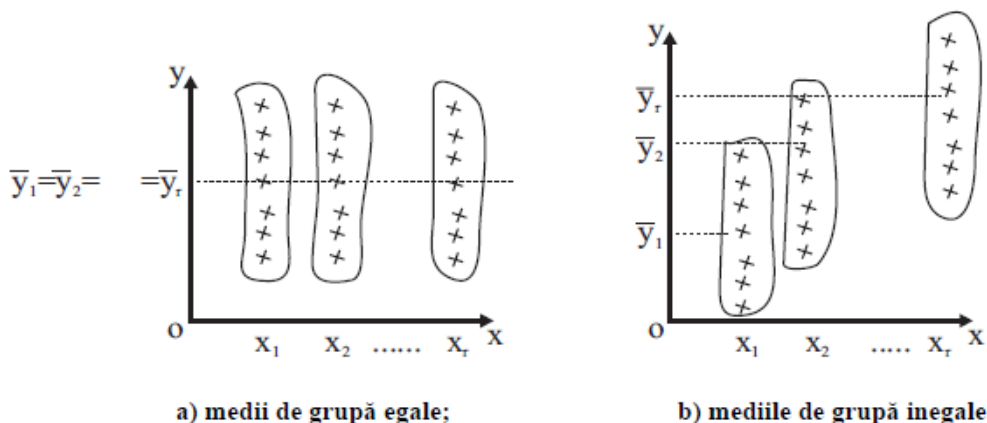
Nu vom ezita și vom aduce câteva exemple de aplicații eficiente ale gândirii statistice atât la nivel instituțional, cât și la nivel individual. Primul exemplu se referă la o adresă venită în anul 1987 la Institutul de Matematică și Informatică al AS din Republica Moldova din partea Poliției arondate la Gara Feroviara din Chișinău. Se cerea ajutor în rezolvarea următoarei probleme. În vara aceluși an a apărut bănuiala că la casele de bilete s-au comis, mai mulți ani la rând, fraude prin intermediul vânzării neautorizate de bilete de călătorie la tren de la aparatele de marcat bilete, aflate în stare de rezerva pentru cazul ieșirii din funcțiune a aparatelor de marcat bilete bază, banii încasați fiind, bineînțeles, sustrași. Poliția de atunci avea, pentru început, la dispoziție doar registrele de vânzare a biletelor pentru aceleași luni de vară din anii precedenți și din anul în curs. În acele registre figurau date privind ziua, luna, anul, casa de marcat bilete, numărul de bilete vândute, dar și numele casierilor. În baza datelor disponibile, problema rezidă în a depista casierii care puteau fi banuiri de fraudare, pentru a nu deranja întreg colectivul de casieri. Primul lucru care se vede din formularea problemei este că se poate aplica Analiza Dispersională, cunoscută și sub numele de analiză de varianță (ANOVA), care a fost introdusă de statisticianul american R.A. Fisher și permite compararea mediilor a două sau mai multe colectivități de date cantitative. Mai exact, s-a aplicat modelul de **analiză dispersională unifactorială** în care se testează ipoteza nulă: *mediile din populații sunt egale*

$$H_0: \mu_{y_1} = \mu_{y_2} = \dots = \mu_{y_r},$$

cu ipoteza alternativă: *cel puțin două medii din populație nu sunt egale*

$$H_1: \mu_{y_i} \neq \mu_{y_j}, (i \neq j),$$

luând în calitate de factor, unul câte unul, fiecare casier, coroborat cu anul și luna în care a emis bilete de călătorie, iar în calitate de populații, biletele vândute de el în anii precedenți și cel curent.



Se testează, cu alte cuvinte, dacă diferențele dintre mediile de grupă din eșantion sunt prea mari pentru a fi atribuite doar întâmplării. Dacă rezultatul testului indică faptul că mediile sunt semnificativ diferite, se concluzionează că factorul X (casierul) are un impact asupra variabilei Y (numărul de bilete vândute). Testul statistic este dezvoltat în concordanță cu următorul raționament. Dacă ipoteza nulă este adevărată, mediile celor  $r$  populații ar trebui să fie, toate, egale. Ne așteptăm atunci ca mediile celor  $r$  eșantioane să fie aproximativ egale. Dacă ipoteza alternativă este adevărată, există diferențe mari între unele medii ale eșantioanelor.

Finalmente, sub incidența banuielilor au cazut 4 casieri. După o jumătate de an de investigații au fost confirmați 3 dintre aceștia, grație descoperirii unor bilete emise de aparatele de marcaj de rezerva în dările de seama ale unor angajați de la instituții din Chișinău care și-au cumpărat bilete, fiind în delegație de serviciu.

Al doilea exemplu pe care l-am ales vine să ilustreze faptul că simpla cunoaștere a *legității probabiliste* care guvernează fenomenul observat și a *Principiului Verosimilității Maxime* este suficientă pentru a trage unele concluzii doar pe baza unei singure observații statistice. Este vorba despre controlul asupra respectării tehnologiilor ce vizează produsele de panificație, în particular producerea cozonacilor cu stafide. Presupunem că pentru producerea a 1000 de cozonaci normele tehnologice prevăd 10000 de stafide și că procesul de amestec a stafidelor în aluat este respectat întocmai. Controlul constă în alegerea la întâmplare a unui cozonac din cei 1000 de cozonaci produși. Admitem că numărul de stafide s-a dovedit a fi egal cu zero! Ce concluzie putem trage din această observație? Mai întâi și întâi de toate ne amintim că, *a priori*, numărul  $X$  de stafide nimerite în cozonacul ales la întâmplare este o variabilă aleatoare (caracteristică statistică) distribuită binomial cu parametrii  $n=10000$  și  $p=1/1000$ , unde  $n$  numărul total al stafidelor iar  $p$  este probabilitatea ca fiecare stafidă concretă din cele 10000 de stafide să nimerescă în cozonacul ales [1]. Prin urmare, probabilitatea teoretică că se va produce evenimentul înregistrat în experimental nostru este egală cu  $P(X=0)=(1-1/1000)^{10000}$ . Numeric, răspunsul calculat în Sistemul Matematica, online, la adresa <https://www.wolframalpha.com> este următorul:

$$\left(1 - \frac{1}{1000}\right)^{10000} = 0.000045173345977048646135460989948496087051066906867668749\dots$$

Dealtfel, dacă nu avem la dispoziție un instrument de calcul numeric performant, putem apela la Teorema Limită a lui Poisson [1], conform căreia

$$P(X=0)=(1-1/1000)^{10000} \approx e^{-10} \approx$$

$$0.000045399929762484851535591515560550610237918088866564969\dots$$

Așadar, s-a produs un eveniment cu o probabilitate foarte mică, ceea ce contrazice *Principiului Verosimilității Maxime* care proclamă că *dacă s-a produs un eveniment aleator, aceasta înseamnă, de regulă, că s-a produs evenimentul cu probabilitatea cea mai mare*. Concluzia ce se impune în acest caz este că, cel mai probabil, tehnologia în cauză nu a fost respectată întocmai.

Or, după cum vedem din cele două exemple aduse de noi, concluziile nu sunt tranșante, ele necesitând niște investigații suplimentare, dar ele confirmă afirmația remarcabilului matematician român, acad. Solomon Marcus care spunea: “Matematica nu este o știință exactă, Matematica este știința aproximărilor cele mai exacte”. Îndrăznim să facem precizarea, că este vorba mai mult despre Matematica Aplicată, chiar dacă matematicienii știu că nu poate fi trasată o graniță absolut distinctă între Matematica Pură și cea Aplicată.

Gradul de aproximare a modelului matematic aplicat în cercetarea statistică și robustețea concluziilor trase din analiza statistică depind, în mare măsură, de specificul fenomenului cercetat. Astfel, putem aduce un exemplu de model (distribuție) probabilistică care contextual se potrivește la descrierea a două experimente aleatoare diferite, dar concluziile depind de specificul acestora. Este vorba de distribuția Hypergeometrică [3] folosită în controlul calității, dar și în sondajele statistice de opinie. În controlul calității se folosește cu succes această distribuție la estimarea statistică a numărului  $M$  de piese ce au defecte ascunse dintr-un număr impunător  $N$  de piese de același tip, produse la aceeași întreprindere. Se extrage, la întâmplare, fără repetare,  $n$  piese, care sunt supuse controlului, aflându-se, astfel, numărul  $k$  de piese defecte,  $0 \leq k \leq \min\{n, M\}$ . Cunoscând valorile  $N$ ,  $n$  și  $k$  metodele statistice ne permit, de exemplu, să estimăm valoarea necunoscută a lui  $M$ , să construim un interval de încredere pentru acest estimator cu probabilitatea de încredere dorită, dar să verificăm ipoteze statistice vizavi de parametrul  $M$ .

Observăm că aceeași schemă de cercetare statistică se aplică și în sondajele de opinie care necesită unul din răspunsurile DA sau NU, sondaj în care se știe numărul întregii populații  $N$ ,

numărul  $M$  a celor care împărtășesc răspunsul DA fiind necunoscut. În baza unui eșantion randomizat de volum  $n$  în care știm numărul  $k$  de răspunsuri DA putem estima statistic valoarea lui  $M$ , etc. Or dacă comparăm aceste două exemple, oricine ne va spune că studiul statistic descris mai sus va da rezultate mai stabile în cazul controlului calității decât cel din sondajul de opinie. De ce? Fiindcă în sondajele de opinie, mai ales cele cu caracter electoral, răspunsurile celor intervievați sunt supuse schimbărilor de moment. Mai mult, respectarea randomizării sondajului este mai dificilă. În plus, însăși procedura de interviu poate influența, așa cum o arată știința Psihologiei. Aceste două exemple ne sugerează că trebuie să manifestăm mai mult realism față de concluziile care le tragem.

## CONCLUZII

Exemplele invocate în lucrarea noastră vin, pe deoparte, să arate eficiența Statisticii ca instrument de gândire inteligentă în economia cunoașterii și, pe de altă parte, că acest instrument trebuie evaluat cu atenție, în funcție de specificul fenomenului sau experimentului aleator la care ne raportăm. Astfel în acele domenii ale Economiei Cunoașterii cum sunt Tehnică, IT, etc., adică acolo unde Principiul Regularității Statistice [4] este ușor verificabil, Statistica oferă rezultate stabile, aceasta spre deosebire de Sociologie, spre exemplu, unde mai trebuie luați în calcul și alți factori, cum ar fi, de exemplu, cei de natură Psihologică.

## BIBLIOGRAFIE

1. A. Leahu, *Probabilități*, Ed. Univ. “Ovidius”, Constanța, 2001, 118 pp.
2. F. Mosteller, R. Rourke, G. Tomas, *Probability: a first course*, Ed. Addison Wesley Publishing Company, Massachusetts, 1961, 426 pp.
3. W. Feller, *An introduction to probability theory and its applications*, V.I, Ed Wiley&Sons, 1981, N.-Y., 498 pp.
4. A. Leahu, P. Chircu, V. Agafitei, *Rolul și locul metodelor probabilistico statistice în economia cunoașterii*, Lucrările CONFERINȚEI ȘTIINȚIFICE INTERNAȚIONALE CONSACRATE CELEI DE-A XXV-A ANIVERSĂRI A ASEM „25 de ani de reformă economică în Republica Moldova: prin inovare și competitivitate spre progres economic” 23-24 septembrie 2016, Vol. VI, Chișinău, pp. 155-157