Cuprins

Lista tabelelor

Lista figurilor

Despre autori

Nicoleta Caragea este conferenţiar universitar la Facultatea de Management Financiar, Universitatea Ecologică din București și expert în cadrul Institutului Naţional de Statistică.

Titlul de doctor în Economie l-a obţinut sub egida Academiei Române, Institutul de Economie Naţională. Activitatea sa didactică se concentrează, în principal, în domeniul statisticii, prin cursuri și seminarii la programele de licenţă și masterat (Statistică, Statistică economică, Statistică socială, Analiză economico-financiară).
În activitatea de cercetare a participat ca expert naţional în diverse proiecte de cercetare ştiinţifică, workshop-uri și conferinţe organizate de instituţii internaţionale de prestigiu cu activitate statistică (EUROSTAT, OCDE, OMS, Banca Mondială, IMF, UNICEF-UIS).
Iniţiatoare a grupului de cercetare R-omanian team, a organizat diferite workshopuri şi conferinţe internaţionale, având ca scop principal introducerea mediului de analiză statistică (R) în statistica oficială din România (http://www.r-project.ro).
Rezultatele activităţii sale de cercetare au fost publicate în numeroase reviste prestigioase din ţară și din străinătate (Revista Română de Economie, Revista Romană de Statistică, Economic Computation and Economic Cybernetics Studies and Research), precum și în baze de date internaţionale recunoscute (Clarivate Analytics, RePEC, Scopus, DOAJ, Index Copernicus, Elsevier etc.).

Scopus Author ID: 56299742400
ResearcherID: C-9002-2018
https://orcid.org/0000-0003-3199-4186

Ciprian Alexandru este conferenţiar universitar la Facultatea de Management Financiar, Universitatea Ecologică din București și expert în cadrul Institutului Naţional de Statistică.

Titlul de doctor în Economie l-a obţinut sub egida Academiei Române, Institutul de Economie Naţională. A participat la un program de studii postdoctorale în care a implementat utilizarea software-ului R ca instrument de analiză a evoluției indicilor bursieri.
Activitatea sa didactică se concentrează, în principal, în domeniul burselor de valori, prin cursuri și seminarii la programele de licență și masterat (Piețe de capital, Managementul Portofoliului, Piețe internaționale de capital).
A participat la diverse proiecte de cercetare, workshopuri, conferințe naționale și internaționale. Activitatea de cercetare a fost pusă în valoare prin publicarea studiilor în reviste din țară și din Europa, precum și în baze de date internaționale recunoscute (RePEC, DOAJ, EBSCO).
În prezent, în cadrul Institului Național de Statistică, participă ca expert în proiecte BigData și utilizează software-ul de analiză statistică R pentru Data cleaning, Data Matching, Web Scraping, analize de date și vizualizare, Data mining, Data integration, data processing, data validation, dar și utilizarea datelor din sursele administrative pentru realizarea de statistici oficiale.

ResearcherID: V-2168-2017
https://orcid.org/0000-0001-8215-6671

Introducere

"...acum nu mai e nimic nou de descoperit;
tot ce rămâne e doar măsurătoarea din ce în ce mai precisă"

— Lord Kelvin (1894)

Cartea tipărită merită răsfoită. Trăim în vremea în care internetul facilitează comunicarea globală, informația fiind disponibilă oricând și oricum. Toată lumea, de la oameni de știință și până la copii de vârstă școlară primesc și oferă informații și propagă idei pe calea internetului. Tirajele publicațiilor, cărților și manualelor tipărite sunt în scădere în întreaga lume, în timp ce postările online captează atenția omenirii.
Obiectivul principal al cărții pe care o propun este de a fi un ghid cuprinzător, în termeni de concepte și tehnici, reprezentativ și, mai ales, practic, în ceea ce privește utilizarea instrumentelor software de analiză statistică, R fiind principalul software utilizat pentru aplicațiile propuse. Ca abordare generală, cartea prezintă principalele concepte utilizate în statistică, cu exemple și explicații descriptive. Exemplele din viața economică - cele mai multe dintre ele bazate pe date statistice reale - problemele rezolvate, dar și cele propuse, acoperă o arie cuprinzătoare de tematici, cititorul având șansa de a fi introdus în sfera aplicativă a conceptelor teoretice parcurse.
Cartea este destinată tuturor celor care doresc să înțeleagă, prin mijloace științifice, fenomenele economice și sociale, sub aspectul măsurării cantitative și din perspectiva determinării cauzale. Deși se adresează, în principal, studenților care se pregătesc să devină specialiști în științele economice, lucrarea este utilă și celor care își propun să cunoască un domeniu atât de frumos și de captivant. Tocmai nevoia de informații, din ce în ce mai complexe, dar și posibilitățile de calcul avansat cu ajutorul soft-urilor tot mai performante, au condus la crearea unui bazin imens de date care pot fi cu ușurință exploatate pe baza analizei statistice. Poate că acesta este și motivul pentru care statistica rămâne o disciplină percepută ca fiind adesea prea matematizată, destinată specialiștilor. Pentru mulți cititori, mai ales dintre cei care nu au o formare bazată pe un aparat matematic, studiul fenomenelor economice prin metode statistice și matematice, presupune un efort

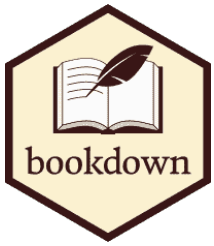
deosebit. Din acest motiv, am încercat să tratez aspectele teoretice, dar și problemele cu aplicație practică din sfera economică, într-o manieră simplă, accesibilă. Așadar, lucrarea are menirea de a facilita înțelegerea conceptelor fundamentale cu care operează statistica, utilizarea adecvată a metodelor de analiză statistică, precum și interpretarea corectă a rezultatelor, în vederea cunoașterii modului de manifestare a fenomenelor.

Nicoleta Caragea
Septembrie, 2018

Bookdown

Această carte a fost editată cu ajutorul pachetului R **bookdown** (Xie, 2015). Cartea are la bază manualul *Statistică - concepte și metode de analiză a datelor*(Caragea, 2015).

Pachetul R **bookdown** este integrat R Markdown (http://rmarkdown.rstudio.com). Documentele elaborate pe baza acestui tip de instrumentar de editare sunt pe deplin reproductibile și dau posibilitatea creării unor formate de ieșire diverse (PDF/HTML/Word/...). Informații suplimentare referitoare la utilizarea pachetului **bookdown** se pot găsi la adresa: https://bookdown.org.



Informații despre software

Software-ul R a devenit în prezent unul dintre cele mai utilizate instrumente de analiză statistică, fiind utilizat în statisticile oficiale, în mediile universitare și de cercetare academică, dar și în mediul de afaceri. Acest manual este destinat tuturor celor care doresc să învețe statistica, fiind un material introductiv de studiu, care prezintă un spectru larg de exemple, prezentări grafice și analiză a datelor, dezvoltate cu ajutorul R.
Aplicațiile din această carte utilizează R, ceea ce înseamnă că pentru reproducerea acestora va fi nevoie de instalarea R pe calculatorul pe care lucrați.
R este un sistem pentru analize statistice și reprezentare grafică creat de către Ross Ihaka și Robert Gentleman, profesori de statistică la Universitatea Auckland din Noua Zeelandă[1].
R este considerat un dialect al limbajului S creat de AT&T Bell Laboratories. S este disponibil sub forma software-ului S-PLUS, comercializat de compania Insightful. Există diferențe importante între cele două limbaje, R și S: acestea sunt documentate de către Ihaka & Gentleman (1996) sau se regăsesc în R-FAQ[2].
Astfel, numele limbajului R provine de la inițiala prenumelui creatorilor, dar este

[1] Ihaka R. & Gentleman R. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299–314.
[2] R-FAQ

totodată și un omagiu adus limbajului S.
În primul rând, R este open-source, fiind distribuit în mod gratuit sub licență *GNU - General Public Licence*[3]; dezvoltarea și distribuirea sunt în grija câtorva profesori și statisticieni, afiliați companiilor și universităților, cunoscuți sub denumirea generică de *R Development Core Team*.
Conform filosofiei *GNU*[4], software-ul open-source este caracterizat de libertatea acordată utilizatorilor săi de a-l utiliza, copia, distribui, studia, modifica și îmbunătăți. Mai exact, este vorba de patru forme de libertate acordate utilizatorilor(Dușa et al., 2015):

- Libertatea de a utiliza programul, în orice scop (libertatea 0);
- Libertatea de a studia modul de funcționare a programului, și de a-l adapta nevoilor proprii (libertatea 1). Accesul la codul-sursă este o precondiție pentru aceasta;
- Libertatea de a redistribui copii, în scopul ajutorării aproapelui tău (libertatea 2);
- Libertatea de a îmbunătăți programul, și de a pune îmbunătățirile la dispoziția publicului, în folosul întregii societăți (libertatea 3). Accesul la codul-sursă este o precondiție pentru aceasta.

Faptul că este gratuit atrage automat avantajul competitiv în fața altor software-uri de analiză statistică, precum Stata, SAS și SPSS. Astfel, costurile alocate licenței de software dispar. R este denumit de către Norman Nie, unul dintre fondatorii SPSS și CEO al Revolution Analytics, "cel mai puternic și flexibil limbaj de programare statistică din lume" (în engleză *"the most powerful and flexible statistical programming language in the world"*).[5] Dovadă a succesului pe care R îl are în știința datelor, s-au dezvoltat medii de integrare a acestuia în SAS și chiar SPSS. Este vorba despre modulul SAS/IML[6], care integrează limbajul R în SAS, și despre *translate2R*, un serviciu de translatare a codului SPSS direct în R dezvoltat de compania *eoda*[7]. R are susținerea comunității științifice, dar și a multor companii internaționale. Dintre acestea, menționăm: Google, Facebook, Mozilla, Twitter, The New York Times, The Economist, NewScientist, Lloyd's, Bing, Johnson&Johnson, Pfizer, Shell, Bank of America, Ford.[8] R este susținut și de mediul academic. Marile universități din lume sprijină R, la fel cum sprijină și alte inițiative sau software-uri open-source, precum sistemul de operare Linux sau sistemul de preparare a documentelor LaTeX.

[3]GNU
[4]GNU Philosophy
[5]Smith, D., 2010,"R is Hot", Revolution Analytics
[6]SAS/IML Module
[7]translate2R - eoda
[8]Revolution Analytics, "Companies Using R"