

Introducere în analiza datelor economico-financiare

Aplicații și exemple rezolvate prin: R, Python, Excel, PowerBI

Ciprian Alexandru-Caragea

9/8/24

Cuprins

Lista figurilor	v
Lista tabelelor	vii
Despre autor	3
Introducere	5
Quarto	7
Informații despre software	7
1 Introducere în analiza datelor	9
1.1 Importanța datelor în luare deciziilor	9
1.2 Identificarea seturilor de date relevante	9
1.3 Concepte de bază legate de formatarea datelor	9
1.4 Instrumente și Software pentru Analiza Datelor	10
2 Introducere în analiza statistică	11
2.1 Concepte de bază privind analiza statistică	11
2.2 Mărimile relative	11
2.3 Sistematizarea datelor	11
2.4 Indicatorii de nivel	11
2.5 Indicatorii variației	11
2.6 Vizualizarea datelor	11
3 Analiza datelor economico-financiare	13
3.1 Indicatori economici și financiari	13
3.2 Principii de bază în economie și finanțe publice	13
3.3 Analiza indicatorilor financiari ai instituțiilor publice	13
4 Analiza bugetară	15
4.1 Structura bugetului public	15

4.2	Metode de evaluare a performanței bugetare	15
4.3	Analiza deviațiilor și optimizarea bugetară	15
5	Etapele analizei datelor	17
5.1	Definirea obiectivelor analizei	17
5.2	Colectarea datelor	17
5.3	Importarea datelor	17
5.4	Stocarea datelor	17
5.5	Curățarea datelor	17
5.6	Agregarea datelor	17
5.7	Analiza descriptivă a seturilor de date	18
5.8	Interpretarea rezultatelor	18
5.9	Vizualizarea rezultatelor	18
5.10	Diseminarea rezultatelor analizei	18
6	Crearea și partajarea rapoartelor și tablourilor de bord	19
6.1	Crearea rapoartelor și tablourilor de bord	19
6.2	Partajarea și colaborarea pe rapoarte	19
7	Calitatea datelor	21
7.1	Standarde, ghiduri și monitorizare	21
7.2	Metodologii	21
7.3	Metadate	21
7.4	Clasificări și nomenclatoare	21
7.5	Rapoarte de calitate	21
7.6	Etica și transparența	21
	Bibliografie	23

Lista figurilor

Lista tabelelor

Despre autor

Ciprian Alexandru-Caragea este conferențiar universitar la Facultatea de Management Financiar, Universitatea Ecologică din București și Analist de Date la diverse instituții internaționale.

Titlul de doctor în Economie l-a obținut sub egida Academiei Române, Institutul de Economie Națională. 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, workshop-uri, 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 Institutului 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 exploatare 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

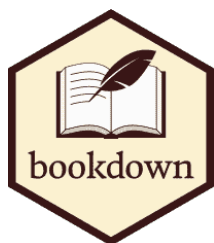
un efort deosebit. Din acest motiv, am încercat să tratez aspectele teoretice, dar și problemele cu aplicație practică din sfera economică, într-o manieră simplă, accesibilă. Așadar, lucrarea are menirea de a facilita înțelegerea conceptelor fundamentale cu care operează statistica, utilizarea adecvată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

Nicoleta Caragea
Septembrie, 2018

Quarto

Această carte a fost editată cu ajutorul pachetului R **bookdown** (xie2015?). Cartea are la bază manualul *Statistică - concepte și metode de analiză a datelor*(caragea2015?).

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 Zeelandă¹.

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-FAQ².

Astfel, numele limbajului R provine de la inițiala prenumelui creatorilor, dar este totodată și un omagiu adus limbajului S.

În primul rând, R este open-source, fiind distribuit în mod gratuit sub licență

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

²R-FAQ

*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(**dusa2015?**):

- 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 condiție 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 condiție 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/IML⁶, care integrează limbajul R în SAS, și despre *translate2R*, un serviciu de traducere 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 L^AT_EX.

³GNU

⁴GNU Philosophy

⁵Smith, D., 2010, “R is Hot”, Revolution Analytics

⁶SAS/IML Module

⁷translate2R - eoda

⁸Revolution Analytics, “Companies Using R”

Capitolul 1

Introducere în analiza datelor

1.1 Importanța datelor în luare deciziilor

text

1.2 Identificarea seturilor de date relevante

text

1.2.1 Surse de date publice și private, inclusiv Open Data

text

1.2.2 Evaluarea calității și relevanței datelor

text

1.2.3 Utilizarea platformelor de date

- Modalități de acces, descărcare și integrare a datelor
- Prezentarea unor platforme comune: Eurostat, INS, UN data, OECD Data

1.3 Concepte de bază legate de formatarea datelor

(formatul și tipurile de date)

1.3.1 Tipuri de surse de date (primare și secundare)

text

1.3.2 Tipuri de fișiere (CSV, XML, JSON) și utilizarea lor

text

1.3.3 Importanța și utilizarea metadatelor

text

1.4 Instrumente și Software pentru Analiza Datelor

1.4.1 Utilizarea Excel pentru analize financiare

text

1.4.2 Utilizarea Power BI pentru analize financiare

text

1.4.3 Introducere în R și Python pentru analize statistice

text

Capitolul 2

Introducere în analiza statistică

2.1 Concepte de bază privind analiza statistică

text

2.2 Mărimile relative

text

2.3 Sistematizarea datelor

text

2.4 Indicatorii de nivel

- medie, mediană, mod, quantile

2.5 Indicatorii variației

- abatere, dispersie, abatere standard, repartiție, asimetrie, concentrare

2.6 Vizualizarea datelor

text

2.6.1 Introducerea statisticii descriptive și inferențiale.

text

2.6.2 Noțiuni fundamentale: medie, mediană, mod, varianță, deviație standard

text

Capitolul 3

Analiza datelor economico-financiare

3.1 Indicatori economici și financiari

text

3.2 Principii de bază în economie și finanțe publice

text

3.3 Analiza indicatorilor financiari ai instituțiilor publice

text

Capitolul 4

Analiza bugetară

4.1 Structura bugetului public

text

4.2 Metode de evaluare a performanței bugetare

text

4.3 Analiza deviațiilor și optimizarea bugetară

text

Capitolul 5

Etapele analizei datelor

- cu exemple practice în Excel, Power BI, R, Python

5.1 Definirea obiectivelor analizei

text

5.2 Colectarea datelor

text

5.3 Importarea datelor

text

5.4 Stocarea datelor

text

5.5 Curățarea datelor

- Tratarea duplicatelor, gestionarea valorilor lipsă

5.6 Agregarea datelor

- Filtrarea/selectarea și agregarea datelor

5.7 Analiza descriptivă a seturilor de date

- Tabele de frecvență, diagrame de distribuție
- Analiza distribuțiilor și identificarea valorile anormale (outliers)

5.8 Interpretarea rezultatelor

text

5.9 Vizualizarea rezultatelor

text

5.10 Diseminarea rezultatelor analizei

text

Capitolul 6

Crearea și partajarea rapoartelor și tablourilor de bord

6.1 Crearea rapoartelor și tablourilor de bord

6.1.1 Dezvoltarea de rapoarte interactive și tablouri de bord în Power BI.

text

6.1.2 Utilizarea storytelling-ului în prezentarea datelor.

text

6.2 Partajarea și colaborarea pe rapoarte

6.2.1 Tehnici de partajare a rapoartelor și tablourilor de bord în organizație

text

6.2.2 Utilizarea platformelor de colaborare (e.g., Google Drive, SharePoint)

text

Capitolul 7

Calitatea datelor

7.1 Standarde, ghiduri și monitorizare

text

7.2 Metodologii

text

7.3 Metadate

text

7.4 Clasificări și nomenclatoare

text

7.5 Rapoarte de calitate

text

7.6 Etica și transparența

text

Bibliografie

