

Decembrie 2017 volum 4
numărul 12 (43)
CUPRINS

Conferința "The Use of R in
Official Statistics" 1-2

Workshop-urile și lectorii
conferinței 3-4

Impresii ale participanților
la Conferința uRos 5

Revizuirea Codului de Bune
Practici al Statisticilor
Europene 6



decembrie 2017
Colectiv redacțional:
Andreea IORDAN
Tania CIMPOERU
Bogdan PĂTĂRLĂGEANU
Cătălin MÎRLOGEANU
Bogdan DRĂGHIA
Hadrian TĂNASE
comunicare@insse.ro
ISSN: 2537 - 3129

6-7 noiembrie 2017

Conferința "The Use of R in Official Statistics"



Conferința Internațională **New Challenges for Statistical Software - The Use of R in Official Statistics - uRos2017** este a 5-a conferință, organizată în România, prin care se promovează utilizarea și dezvoltarea soft-ului de analiză R în statistica oficială din România, dar și a soft-urilor care integrează pachete și funcții R. Conferința Internațională uRos are ca principal organizator din INS pe Nicoleta Caragea, conferențiar la Facultatea de Economie a Universității Ecologice și expert superior la Institutul Național de Statistică, Direcția de calcul a



indicatorilor privind populația și migrația externă. Nicoleta Caragea a participat ca expert național în diverse proiecte, workshop-uri și conferințe organizate de instituții internaționale de

prestigiu cu activitate statistică (EUROSTAT, OECD, WHO, World Bank, UNICEF-UIS).

Scopul conferinței este de a oferi un forum public pentru statisticieni și experți din institutele de statistică din toate statele lumii, dar și pentru cercetători din mediul academic. Conferința oferă oportunitatea de a face schimb de idei și de a dezbate evoluția mediului de analiză statistică R.

Conferința s-a desfășurat în perioada 6-7 noiembrie 2017 și a reunit peste 60 de participanți statisticieni și specialiști din oficiile de statistică și din mediul academic din 20 de țări (Austria, Canada, Columbia, Croația, Franța, Germania, Italia, Irak, Japonia, Lituania, Luxemburg, Maroc, Olanda, Norvegia, Polonia, România, Spania, Elveția și Turcia). De asemenea, au fost prezenți reprezentanți ai Eurostat și ai altor organizații internaționale care au ca obiect de activitate statisticile oficiale (Națiunile Unite/UNIDO și FAO).

continuare în pagina 2

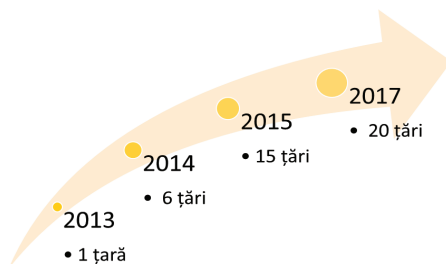
Conferința "The Use of R in Official Statistics"



urmare din pagina 1

Lucrările conferinței s-au desfășurat pe secțiuni paralele și au cuprins 22 de prezentări și 8 workshop-uri tematice.

De-a lungul celor cinci ediții ale evenimentului, participarea internațională a crescut exponențial.



Evenimentul de anul acesta a găzduit, pe lângă sesiunile de prezentări, 8 workshop-uri ale unor lectori de prestigiu din statistica oficială și mediul academic: Mark van

der Loo, Valentin Todorov, Bernhard Meindl, Marcello D'Orazio, Camelia Goga, Hervé Cardot, Bogdan Oancea, Ciprian Alexandru, Elena Druică

Prezentările Conferinței uRos2017 sunt disponibile la adresa web:

<http://www.r-project.ro/conference2017/presentations.html>

Lucrările conferinței se publică în două numere ale Revistei Române de Statistică: numărul 4/2017 și numărul 1/2018. Primul a fost deja publicat și înmănat participanților în timpul conferinței, iar cel de-al doilea va fi tipărit în martie 2018.

Calitatea lucrărilor și a workshop-urilor prezentate, bogata activitate desfășurată de autorii cercetărilor, dar și rezultatele obținute pot constitui tot atâtea puncte forte, de referință, pentru a concluziona că România este prezentă pe harta utilizatorilor mediului de analiză statistică R.

Conferința a constituit o oportunitate de a dezvolta noi idei și cooperări în domeniul statisticii oficiale și a mediului academic. Acest eveniment a adus în prim plan rolul Institutului Național de Statistică în statistica oficială.

O altă noutate a acestei ediții a conferinței din România este faptul că evenimentului i s-a alăturat, sub titulatura de Joint Conference, Conferința Internațională desfășurată în Malaysia, "International Conference On Computing, Mathematics And Statistics 2017" (iCMS2017) <http://www.icmskdh.net/>. În cadrul conferinței uRos 2017 Asia Pacific/ iCMS2017 au participat, în calitate de speakeri, Nicolaas Jan Dirk Nagelkerke, Matthias Templ și Martin Everett.

Website-ul conferinței uRos2017: <http://r-project.ro/conference2017/>

În cadrul Conferinței Internaționale uRos 2017 a avut loc un eveniment conex în data de 8 noiembrie 2017 – o întâlnire între delegația din Japonia și Austria (Națiunile Unite/UNIDO) și reprezentanți ai Institutului Național de Statistică, în vederea schimbului de cunoștințe din domeniu. Subiectele abordate în cadrul acestei întâlniri au fost:

- Modernizarea statisticii oficiale din România
- Utilizarea R în anchetele statistice
- Editarea de date (identificarea outliers-ilor, imputare etc.)



- Generarea de rapoarte statistice utilizând pachetele Sweave și knitr ale programului R

- Colectarea de date online în cadrul anchetelor din statistica întreprinderii

Workshop-urile și lectorii conferinței

Olanda: **Mark van der Loo** - *Statistics Netherlands*

Validarea datelor statistice cu ajutorul R

Mark van der Loo este autorul și coautorul mai multor lucrări și pachete R care sunt în special legate de metodologia de validare a datelor. Acestea includ pachetele R „stringdist”, „validate”, „deductive” și „lintools”. Este consultant intern, cercetător științific și responsabil pentru distribuția R la *Statistics Netherlands*. De asemenea, predă despre

R și metodologia de validare a datelor, precum și un curs de management al datelor la Universitatea Utrecht pentru programul EMOS (European Master of Official Statistics).



Franța: **Camelia Goga** - *Institut de Mathématiques de Bourgogne, Université de Bourgogne*

Tehnici de eșantionare cu R

Membră a echipei SPOC (Statistică, Probabilități, Optimizare și Control) a Institutului de Matematică



din Bourgogne, Camelia Goga este profesor universitar de statistică la Universitatea Franche-Comté din Bourgogne. Domeniile sale de cercetare includ tehnicile de eșantionare a sondajelor și statisticile nonparametrice și funcționale.

<http://goga.perso.math.cnrs.fr/>

Austria: **Bernhard Meindl** - *Statistics Austria*

Validarea datelor statistice cu ajutorul R

În cadrul workshop-ului au fost evidențiate evoluțiile actuale ale instrumentelor R privind confidențialitatea statistică. sdcMicro (pentru anonimizarea microdatelor) și sdcTable (pentru protejarea datelor tabelare) pot fi utilizate în acest domeniu. Principalele evoluții actuale sunt reprezentate de o nouă interfață

grafică pentru sdcMicro, pe care am arătat-o publicului într-o sesiune live. Pentru sdcTable s-au făcut unele modificări pentru a folosi tau-argus ca o soluție de rezervă.



Italia: **Marcello D’Orazio** *Organizația Națiunilor Unite pentru Alimentație și Agricultură*

Detectarea valorilor extreme cu R: câteva remarci

Marcello D’Orazio, doctor în statistică la Universitatea din Bari (Italia), a început să lucreze la Institutul Național de Statistică din Italia (Istat) în 1998 ca cercetător în metodologia statistică (cercetător principal din 2008). La Istat a fost implicat în diverse proiecte privind calitatea statisticilor oficiale, prelucrarea datelor din cercetări statistice, statisticile agricole, utilizarea R în producerea de statistici oficiale. Din aprilie 2016 este detașat la Organizația pentru Alimentație și Agricultură (FAO) a Națiunilor Unite, oferind expertiză metodologică și de management în sprijinul activității Directorului și Diviziei de Statistică.



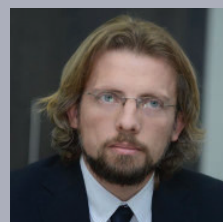
Workshop-ul „Detectarea valorilor extreme cu R: câteva remarci” a oferit o imagine de ansamblu asupra metodelor de bază și a celor avansate pentru a detecta valori extreme în datele anchetei sau, la modul general, alte surse de date care pot fi utilizate pentru producerea statisticilor oficiale. De asemenea prezentarea a arătat că R este un produs dezvoltat îndeajuns pentru realizarea unei astfel de sarcini, având în vedere că multe funcții sunt deja disponibile în diverse pachete accesibile gratuit pe CRAN. Prezentarea a fost axată în principal pe valori univariate și bivariate și, în special, oferă câteva informații despre modul de abordare a datelor când se compară valorile actuale cu cele observate în trecut.

România: **Ciprian Alexandru-Caragea**

Universitatea Ecologică București, Institutul Național de Statistică

De la date nestructurate la date structurate – extragerea datelor de pe site-uri Web pentru statisticile oficiale

Ciprian Alexandru este conferențiar universitar și prorector al Universității Ecologice din București. Este autorizat ca și consultant de investiții pe piața de capital din România. Doctoratul său în Economie a fost obținut sub egida Academiei Române, Institutul Național de Cercetări Economice. A finalizat un program postdoctoral în care a pus în aplicare utilizarea lui R ca instrument de analiză a evoluției indicilor pe piața de capital.



Bulgaria: Valentin Todorov - Organizația Națiunilor Unite pentru Dezvoltare Industrială (UNIDO)

R în biroul de statistică: experiența UNIDO

Valentin Todorov este un ofițer senior în gestionarea datelor la Organizația Națiunilor Unite pentru Dezvoltare Industrială (UNIDO). A obținut o diplomă de doctor în statistică la Universitatea de Tehnologie din Viena, Austria și deține o diplomă de masterat în matematică și statistică la Universitatea din Sofia, Bulgaria și Universitatea din Viena. Activitățile sale principale de cercetare includ analiza datelor multivariate, statisticile computaționale, robust statistics, statisticile oficiale și sistemele informatice statistice. Valentin Todorov a scris o carte, diferite capitole în alte lucrări și multe alte articole științifice în domeniul sta-

tisticilor matematice și sistemelor informatice statistice. A scris de asemenea recenzii pentru o serie de reviste internaționale de statistică. A dezvoltat și menținut în cadrul CRAN mai multe pachete R. Înainte de a se alătura UNIDO în 2006, Valentin Todorov a lucrat în proiecte internaționale de anvergură în domeniile aviației (Austrocontrol), telecomunicațiilor (Ericsson) și academice (Academia Bulgară de Științe).



România: Bogdan Oancea
Institutul National de Statistica

De la date nestructurate la date structurate - utilizarea tehnicilor web-scraping pentru producerea unor statistici experimentale

Bogdan Oancea este director al Direcției de instrumente inovatoare în statistică. El este specialist în analiza datelor, limbaje de programare, calcul științific și econometrie. Bogdan Oancea are experiență în dezvoltarea produselor software, participând la diferite proiecte dezvoltate folosind limbajele de programare C, C++, Java sau R precum și tehnologii specifice domeniului big-data precum Hadoop, MPI, OpenMP, CUDA. Bogdan Oancea este membru al ESS Big Data Task-Force.

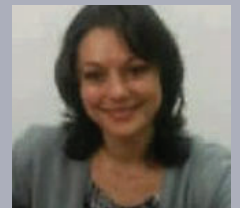
România: Elena Druică

Universitatea din București, Facultatea de Științe Economice și Administrative

Aplicație a pachetului „pglm” din R. Analiza numărului de infecții nosocomiale în spitalele din România

Dr. Elena Druică este profesor universitar de Econometrie Aplicată la Facultatea de Afaceri și Administrație, Universitatea din București. A obținut o diplomă în matematică de la Universitatea din București în 1994, o diplomă în Analiza Armonică în 1995 și o diplomă de master în Managementul Financiar și al Riscurilor în 2004. A obținut o diplomă de doctorat

în Statistică și Probabilități în cadrul Academiei Române de Științe în 2004, și un doctorat în Economie la Academia de Studii Economice în 2006. Este membră a echipei R din România.



Franța: Hervé Cardot - Institut de Mathématiques de Bourgogne, Université de Bourgogne

Estimări rapide și robuste, clusterizarea și Analiza Principală a Componentelor folosind eșantioane de dimensiuni mari în R

Profesor la Universitatea din Bourgogne, membru al echipei SPOC (Statistică, Probabilități, Optimizare și Control) a Institutului de Matematică din Bourgogne. Hervé Cardot a devenit doctor în statistică în 1997 la Universitatea Paul Sabatier (Toulouse, Franța). Este profesor de statistică la Universitatea din Bourgogne, Dijon, din 2007 și șef al echipei SPOC (Statistică, Probabilități, Optimizare și Control) la Institutul de Matematică din Bourgogne.

Domeniile sale de cercetare sunt analiza funcțională a datelor, claritatea și robustețea eșantioanelor de dimensiuni mari, eșantionarea, precum și aplicarea statisticilor în diverse domenii (climatologie, economie, proteomică, energie, teledetecție, senzometrie,...)

<http://cardot.perso.math.cnrs.fr/>



Impresii ale participanților la Conferința uRos

Mark van der Loo, CBS Olanda

- Cum ați descrie experiența utilizării software-ului R în munca dumneavoastră zilnică? Cum este R diferit de software-urile similare pe care le-ați folosit?

Foarte simplu: Nu pot să lucrez fără R. Productivitatea mea ca metodolog ar scădea semnificativ dacă aș fi forțat să lucrez cu software de tip point-and-click, cum ar fi Excel sau SPSS. Probabil că aș fi „supraviețuit” dacă am fi folosit Python în loc de R, cu excepția faptului că Rstudio este mult mai bun decât orice interfață construită pentru Python. De asemenea, rmarkdown este de foarte mare ajutor!

- Acum, în 2017, la sediul INS România s-a desfășurat pentru a cincea oară Conferința uRos. În toți acești ani, numărul utilizatorilor de software a

crescut permanent. De ce credeți că statistica oficială din România este din nou gazda acestei conferințe și ce părere aveți despre viitorul software-ului statistic?

- Nu sunt sigur de ce România a fost gazdă din nou, dar sunt foarte recunoscător că tot dumneavoastră ați organizat conferința! Aceste întâlniri sunt importante și cred că această conferință va beneficia de pe urma faptului că este ținută în mai multe locuri, astfel încât să poată avea ca public cât mai mulți utilizatori.

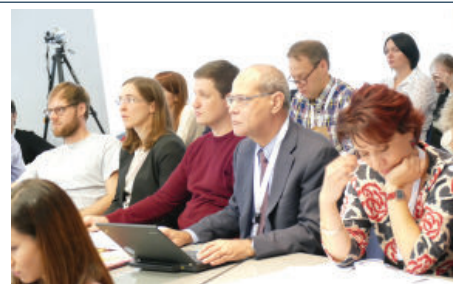
- Cum descrieți informațiile pe care le-ați obținut în aceste două zile de participare intensă la conferința din București?

Am crezut că este un moment de inspirație să aud că în toate oficiile

naționale de statistică, introducerea R începe cu acțiunea unui „lup singuratic”: un entuziast rebel care a introdus R „neoficial” și totul începe de acolo. Cred că este foarte important să împărtășim aceste experiențe pentru a stimula oamenii care sunt pionierii R ai institutului lor. De asemenea, am aflat de câteva pachete noi interesante, în special în cadrul workshop-urilor.

- Vă rugăm spuneți-ne câteva cuvinte despre participarea dumneavoastră la Conferința R (gazde, locație etc.)

- Gazdele au fost foarte amabile și de mare ajutor ca de fiecare dată. Îmi place atmosfera informală, iar programul social a fost foarte frumos de asemenea. Am petrecut un timp minunat și aștept cu nerăbdare ocazia de a organiza uRos 2018!



Bernhard Meindl, Statistics Austria

- Cum ați descrie experiența utilizării software-ului R în munca dumneavoastră zilnică? Cum este R diferit de software-urile similare pe care le-ați folosit?

- Eu folosesc R exclusiv în munca mea zilnică. Îndeplinesc sarcini legate de confidențialitatea datelor, de raportare automată, crearea de grafice și aplicații interactive, care pot face munca mai ușoară pentru unii colegi. Deoarece R este accesibil și gratuit, este mult mai ușoară partajarea instrumentelor și codificarea împreună cu colegii, ceea ce este adesea dificil cu produsele software protejate de proprietatea intelectuală.

- Acum, în 2017, la sediul INS România s-a desfășurat pentru

a cincea oară Conferința uRos. În toți acești ani, numărul utilizatorilor de software a crescut permanent. De ce credeți că statistica oficială din România este din nou gazda acestei conferințe și ce părere aveți despre viitorul software-ului statistic?

- R are cu siguranță un viitor strălucit. Numărul de pachete software R este în creștere și sunt convins că R are un loc în aproape orice domeniu în care Institutele de Statistică sunt implicate, de la gestionarea datelor până la validarea datelor, eșantionarea, confidențialitatea datelor, aplicațiile interactive, extragerea datelor de pe site-uri Web, problemele de big data, gestionarea metadatelor și multe altele.

- Cum descrieți informațiile pe care le-ați obținut în aceste două zile de participare intensă la conferința din București?

- Este minunat să aflăm că utilizarea R în statistica oficială crește de la an la an și să vedem cum se înregistrează progrese în multe țări. A fost interesant să aud că, de exemplu, primul pachet R din România tocmai a fost prezentat!

- În cele din urmă, spuneți-ne câteva cuvinte despre participarea dumneavoastră la Conferința R (gazde, locație etc.)

- Gazdele au fost foarte amabile, locația a fost potrivită și mâncarea bună. Cina care a urmat conferinței a fost, de asemenea, foarte plăcută și mi-a făcut cu adevărat plăcere să particip la această conferință.

În Buletinul informativ al INS din luna ianuarie vă vom prezenta și alte impresii ale participanților la Conferința „The Use of R in Official Statistics”

Revizuirea Codului de Bune Practici al Statisticilor Europene

La data de 9 octombrie 2017 a avut loc la Luxemburg cea de-a 21-a întâlnire a Grupului de Lucru privind Calitatea în Statistică. Cu această ocazie, Eduardo Barredo Capelot, directorul Departamentului pentru asigurarea calității din cadrul Eurostat cel ce a prezidat reuniunea, a explicat participanților că în prezent prioritatea pentru Eurostat în ceea ce privește evoluția calității o constituie revizuirea Codului de Bune Practici al Statisticilor Europene (CoP). Din acest motiv, consultarea planificată inițial a fost înlocuită cu această întâlnire propriu zisă, pentru a da posibilitatea reprezentanților din Sistemul Statistic European să își exprime ideile și opiniile legate de acest subiect și să poată dezbate și lua decizii cu privire la diferitele întrebări și comentarii tehnice care au fost ridicate în timpul schimbului de opinii de la reuniunea Comitetului SSE din septembrie 2017.

Vă prezentăm propunerile pentru modificarea Codului de Bune Practici al Statisticilor Europene.

Această ediție din 2017 este cea de-a doua revizuire a Codului european de practici în statistică, adoptat inițial în 2005. Scopul revizuirii este de a reflecta ultimele schimbări și inovații în dezvoltarea, producerea și difuzarea statisticilor oficiale în Sistemul Statistic European și în afara acestuia, sursele, utilizarea noilor tehnologii, modernizarea cadrului juridic și rezultatele evaluărilor Peer Review privind punerea în aplicare a Codului de bune practici.

Principiile relevante sunt independența profesională, coordonarea și cooperarea, mandatul pentru colectarea datelor, gradul de adecvare a resurselor, angajamentul de calitate, confidențialitatea statistică, imparțialitatea și obiectivitatea.

Alte modificări și precizări convenite: În ceea ce privește indicatorul 1.8, *procedurile de recrutare* și numire a șefilor instituțiilor naționale de statistică și ale Eurostat și, după caz, șefilor statisticieni ai altor autorități statistice, *trebuie să fie transparente și să se bazeze numai pe criterii profesionale*.

A fost introdus **Principiul 1bis: Coordonare și cooperare**. Institutele naționale de statistică și Eurostat asigură coordonarea tuturor activităților de elaborare, producție și diseminare a statisticilor europene la nivelul sistemului statistic național și, respectiv, al sistemului statistic european. Autoritățile statistice cooperează activ în cadrul parteneriatului Sistemului Statistic European, pentru a asigura elaborarea, producerea și diseminarea statisticilor europene.

Indicatorul 1bis.1: Institutele naționale de statistică coordonează activitățile statistice ale tuturor celorlalte autorități naționale care elaborează, produc și difuzează statistici europene. *Aceștia acționează în acest*

sens ca unicul punct de contact al Eurostat în probleme statistice. Legislația și procedurile bine definite și stabilite sunt gata pentru punerea în aplicare a rolului de coordonare la nivel național și european.

Indicatorul 1bis.2: Orientările naționale pentru asigurarea calității în dezvoltarea, producerea și diseminarea statisticilor europene în cadrul sistemului statistic național sunt elaborate de șefii instituțiilor naționale de statistică, dacă este necesar; implementarea acestora este monitorizată și revizuită.

Indicatorul 1bis.3: Autoritățile statistice mențin și dezvoltă continuu cooperarea la diferite niveluri între ele și cu organele consultative ale Sistemului Statistic European, precum și cu membrii Sistemului European al Bancilor Centrale, ai instituțiilor academice și ai altor organisme internaționale.

Principiul 2: Mandatul pentru culegerea datelor și accesul la date. Autoritățile statistice dețin un mandat legal clar pentru culegerea și accesarea informațiilor *din mai multe surse de date* în vederea elaborării de Statistici Europene. Administrațiile, întreprinderile, gospodăriile și publicul larg pot fi obligați prin lege să permită accesul sau să furnizeze date pentru elaborarea Statisticilor Europene, la cererea autorităților statistice.

Indicatorul 2.1: Mandatul autorităților statistice pentru culegerea de informații și accesul la informații *din mai multe surse de date* în vederea dezvoltării, producerii și diseminării statisticilor europene este stipulat în lege.

Indicatorul 2.2: Autoritățile statistice sunt autorizate prin lege să acceseze datele administrative, în mod prompt și gratuit, și să le utilizeze în scopuri

statistice. *Acestea sunt implicate încă de la început în conceperea, dezvoltarea și întreruperea înregistrărilor administrative, pentru a le face mai potrivite pentru scopuri statistice.*

Indicatorul 2.4: Accesul în scopuri statistice la alte date, cum ar fi datele private, este facilitat, asigurând în același timp confidențialitatea statistică și protecția datelor.

Principiul 3: Adecvarea resurselor. Resursele de care dispun autoritățile statistice sunt suficiente pentru a răspunde cerințelor Statisticilor Europene.

Indicatorul 3.1: Sunt disponibile resurse *umane*, financiare și *tehnice*, adecvate atât din punct de vedere cantitativ cât și calitativ, pentru a răspunde cerințelor curente de statistici.

Principiul 4: Angajamentul privind calitatea. Autoritățile statistice se angajează să asigure calitatea. Acestea identifică în mod sistematic și regulat punctele tari și punctele slabe, pentru îmbunătățirea continuă a calității proceselor și *produselor statistice*.

Indicatorul 4.2: Există proceduri de planificare, monitorizare și îmbunătățire a calității proceselor statistice, *inclusiv integrarea datelor din surse multiple de date*.

Indicatorul 4.3: *Calitatea produselor statistice* este monitorizată regulat, evaluată în ceea ce privește eventualele soluții de compromis și raportată în conformitate cu criteriile de calitate ale Statisticilor Europene.

În Buletinul informativ al INS din luna ianuarie vă vom prezenta și propunerile de modificare a celorlalte principii din cadrul Codului de Bune Practici al Statisticilor Europene