

Indicatii de analizare a datelor folosind R cu interfata R Commander

Cuprins

Instalare R cu interfata R Commander	2
Instalarea simpla – recomandata !!!	2
Lansarea programului	2
Instalarea clasica	2
Lansarea R – fara interfata	3
Cum se instalează interfața grafică R Commander	3
Pornirea interfetei grafice R Commander (Rcmdr)	4
Configurarea R să pornească automat cu interfața grafică R Commander	4
Configurarea Rcmdr să pornească automat cu anumite plug-inuri încărcate	4
Instalarea pe sisteme Mac OS sau Linux	5
Importarea fișierelor Excel in R Commander	5
Modificarea variabilelor din setul de date	6
Schimbarea ordinii categoriilor unei variabile calitative	6
Descrierea numerica a variabilelor	6
Calitative	6
Cantitative	6
Teste statistice	7
Comparare date calitative / relatii între date calitative	7
2/>2 esantioane independente	7
2 esantioane dependente/perechi	7
Comparare date cantitative	8
2 esantioane independente	8
2 esantioane dependente/perechi	8
Comparare date de supravietuire	8
Calculul indicatorilor medicali: OR, RR, RA, Se, Sp, VPP, ARR, NNT,	9
Corelatii	9
Calculul coeficientilor de corelatie (Pearson/Speaman) si valoarea lui p a testului asociat lor	9
Regresii lineare	10
Semnificația statistica a modelelor (p corespunzator lui F significance), coeficienții de de determinare, precum și tabelele cu coeficienții variabilelor	10
Regresii pentru date de supravietuire	10
Regresia Cox – aflarea ratei hazardului	10

Realizarea tabelelor	11
Tabele de frecventa	11
Tabele de contingenta.....	11
Reprezentarea grafica a rezultatelor	11
Optiuni utile pentru grafice (marime text, culori, teme)	11
Pentru variabile calitative / cantitative discrete.....	12
Descriere variabile cantitative	12
Evaluare legatura intre 2 variabile calitative.....	12
Pentru variabile cantitative.....	12
Descriere date cantitative	12
Comparare date cantitative intre grupuri (in functie de o variabila de grupare)	13
Evaluare legatura intre doua variabile cantitative	13
Pentru date de supravietuire	13
Descriere date supravietuire	13
Copierea / salvarea graficelor din R Commander in alte programe.....	14

Instalare R cu interfata R Commander

Instalarea simpla – recomandata !!!

Se copiaza de pe site (pagina cu laboratoarele – sus) arhiva programului R preinstalat pe calculator. Se dezarchiveaza programul intr-un loc pe hard-diskul propriu (ex. Pe partitia D:).

Lansarea programului

Se intra in R-2.15.1/bin/i386 si se porneste programul Rgui.exe. Atentie daca aveti un calculator cu procesor pe 64 de biti intrati in R-2.15.1/bin/x64 si porniti programul Rgui.exe.

Puteti sa va faceti un shortcut pentru acest program pe Desktop, pentru un acces usor ulterior. Click dreapta pe Rgui.exe – in meniul aparut selectati Send to – Desktop (create shortcut). Pe Desktop va apare o icoana cu un R albastru pe care o puteti folosi sa porniti programul cu usurinta.

Instalarea clasica

Aceasta se adreseaza celor care se pricep mai bine sa lucreze pe calculator.

Se accesează site-ul <http://www.r-project.org/> (The R Project for Statistical Computing). Chiar pe prima pagină există secțiunea Getting Started cu un link **download R** unde se dă click. Sau se poate selecta din meniul din stânga, secțiunea Download, Packages link-ul **CRAN**. În pagina deschisă (CRAN Mirrors) trebuie selectat un server de unde să fie downloadat R (ex. Germania, Franța, UK ...). Se dă click pe oricare server (ideal cât mai aproape geografic de România, pentru ca viteza de download să fie bună și serverele să nu fie solicitate). În noua pagină deschisă **The Comprehensive R Archive Network**, secțiunea de sus din pagină, **Download and Install R**, se selectează link-ul pentru **Windows**. Apare o nouă pagină, R for Windows, unde se urmează legătura numită **base** (binaries for base

distribution). Apare o nouă pagină, intitulată **R-numărul versiunii for Windows (32 bit build)** (ex. *R-2.15.1 for Windows*). Se urmează link-ul din susul paginii **Download R numărul versiunii for Windows** (ex. *Download R 2.15.1 for Windows*). Browserul cu care navigați vă anunță că **puteți downloada** fișierul de instalare al programului R. **Salvați-l** undeva convenabil pe hard disk. Căutați fișierul salvat în locul unde l-ați salvat pe hard disk. Fișierul se numește R-numărul versiunii-win32.exe (ex. *R-2.15.1-win32.exe*). **Deschideți** fișierul, iar dacă Windows-ul vă întreabă "The publisher could not be verified. Are you sure you want to run this software?", apăsați butonul **Run**.

Apare fereastra de mai jos, în care selectați limba în care doriți să faceți instalarea (ex. English) și apoi se apasă butonul **OK**

În noua fereastră, afișată mai jos, se apasă butonul **Next**

Apare o fereastră prezentând licența de soft gratuit GPL și se apasă butonul **Next**

În noua fereastră apărută se lasă calea de instalare sugerată de programul de instalare, și se apasă butonul **Next**:

În fereastra următoare sunteți întrebați ce componente doriți să instalați. Lăsați selecția făcută în mod implicit, eventual bifați PDF help pages (reference manual), iar apoi apăsați butonul **Next**:

În fereastra apărută vă întreabă dacă doriți să modificați modul de pornire a programului R. Selectați **Yes (customized startup)** ! și apoi apăsați butonul **Next**.

În fereastra nou deschisă suntem întrebați de modul de afișare. **Selectați SDI (separate windows)**, iar apoi apăsați butonul **Next**.

În fereastra care vine, suntem întrebați de modul în care dorim să fie afișat Help-ul (informațiile de ajutor). Selectați **HTML help** și apăsați butonul **Next**.

Fereastra nou deschisă ne întreabă de modul de acces la internet. Dacă nu folosiți proxy, sau nu știți ce să alegeți utilizați opțiunea **Standard**, și apăsați butonul **Next**.

Următoarea fereastră ne întreabă cum să se numească folderul din meniul Start de unde va putea fi găsită și lansată aplicația R. Se lasă R în câmpul de editare și se apasă butonul **Next**.

O nouă fereastră ne întreabă despre ultimele opțiuni. Lăsați așa cum sunt selectate, eventual bifați în plus Create a Quick Launch icon, dacă folosiți frecvent programul R, iar apoi apăsați butonul **Next**.

Fereastra următoare arată progresul instalării programului R.

În continuare apare fereastra care ne informează că R a fost instalat pe calculator. Se apasă butonul **Finish**.

Lansarea R – fara interfata

După ce programul R a fost instalat, el se lansează apăsând butonul **Start**. Se deschide **Start Menu**, unde există folderul **R**, în care se află programul R-numărul versiunii (ex. R 2.15.1), pe care se dă click. Apare fereastra în linie de comandă a programului R:

Cum se instalează interfața grafică R Commander

Înainte de instalare a oricărui package (pachet), trebuie să vă asigurați că aveți o conexiune la internet funcțională.

Se pornește R. În fereastra **R Console**, se accesează **meniul Packages**, comanda **Install package(s)...**

Apare o fereastră: **CRAN mirror**, în care se **selectează** un **server** cât mai apropiat de România, ex. Germania și se dă click pe butonul **OK**.

Noua fereastră apărută, se numește **Packages** și în ea sunt afișate toate package-urile disponibile pe serverele CRAN. Spre exemplificare instalăm package-ul Rcmdr care oferă o

interfață grafică utilă. Pentru aceasta **selectăm** Rcmdr din lista de package-uri și apăsăm butonul **OK**.

În cazul în care apare fereastra următoare (Question: Would you like to create a personal library ... to install packages into?), se dă click pe butonul Yes.

Întrucât package-urile instalate pot să depindă de alte package-uri, este bine să încercăm să le încărcăm pentru a downloada cu această ocazie și package-urile de care depind.

Pentru aceasta în fereastra **R Console**, se scrie comanda library(nume package), de ex: library(Rcmdr). Atenție R ține cont dacă literele sunt scrise mic sau mare, și le consideră diferite.

În urma pornirii pachetului Rcmdr, apare o fereastră care ne informează ce pachete lipsesc și ne întreabă dacă să le instalăm. Dăm click pe butonul Yes.

Apare o nouă fereastră **Instalare pachete lipsă**, în care ne întrebă de unde să instalăm pachetele. Selectăm CRAN și apăsăm butonul OK.

R downloadază de pe Internet și instalează apoi toate pachetele lipsă. Trebuie așteptat câteva minute pentru ca operațiunea să se finalizeze.

Pornirea interfeței grafice R Commander (Rcmdr)

În primul rând trebuie ca package-ul Rcmdr să fie instalat, pentru aceasta citiți capitolul "Cum se instalează un package în R".

Pentru a porni interfața grafică Rcmdr, deschideți R, iar în fereastra **R Console**, se scrie comanda library(Rcmdr). Atenție R ține cont dacă literele sunt scrise mic sau mare, și le consideră diferite.

În urma acestei comenzi apare fereastra de lucru grafic Rcmdr:

Configurarea R să pornească automat cu interfața grafică R Commander

Dacă doriți ca pornind R să pornească automat cu interfața grafică Rcmdr, atunci căutați fișierul Rprofile.site. Pentru a-l găsi mergeți în folderul unde ați instalat R-ul. Dacă nu vă aduceți aminte unde s-a instalat, cel mai probabil este pe partiția C:, în folderul Program Files. Acolo veți vedea un folder R. În interiorul acestui folder se află versiunea de R instalată de dumneavoastră ca și folder ex: R-2.15.1. Intrați în acest folder și veți vedea folderul etc. În folderul etc veți vedea fișierul Rprofile.site. Pentru a-l edita, deschideți programul Notepad (pe care îl găsiți apăsând butonul de Start, în All Programs, meniul, Accessories – Notepad). Trageți fișierul Rprofile.site peste fereastra de Notepad pentru a-l deschide. Sau tot pentru a-l deschide puteți din meniul File al Notepad, comanda Open să căutați fișierul pe hard disk și să-l deschideți.

După ce ați deschis fișierul Rprofile.site, copiați următoarele linii de cod în fișier, în continuarea conținutului deja existent ^[1]:

```
local({
old <- getOption("defaultPackages")
options(defaultPackages = c(old, "Rcmdr"))
})
```

Salvați modificările, selectând în meniul **File** comanda **Save**.

Configurarea Rcmdr să pornească automat cu anumite plug-inuri încărcate

Rcmdr poate fi extins prin plug-inuri. Un plug-in activat va putea fi utilizat până la închiderea Rcmdr. La o nouă repornire el trebuie reactivat. Pentru ca Rcmdr să se pornească cu anumite plug-inuri în mod automat trebuie adăugate plug-inurile preferate în fișierul Rprofile.site (vezi capitolul: Cum se configurează R să pornească automat cu interfața grafică Rcmdr?).

Față de opțiunile de pornire automată a Rcmdr se adaugă și plug-inurile preferate sub formă de listă. De exemplu în configurația fișierului Rprofile.site de mai jos au fost precizate plug-inurile RcmdrPlugin.survival și RcmdrPlugin.epack să pornească automat cu Rcmdr. Se pot adăuga unul sau mai multe astfel de plug-inuri, scrise între ghilimele și despărțite prin virgule.

```
local({  
old <- getOption("defaultPackages")  
options(defaultPackages = c(old, "Rcmdr"), Rcmdr=list(plugins=c("RcmdrPlugin.survival",  
"RcmdrPlugin.epack")))  
})
```

Instalarea pe sisteme Mac OS sau Linux

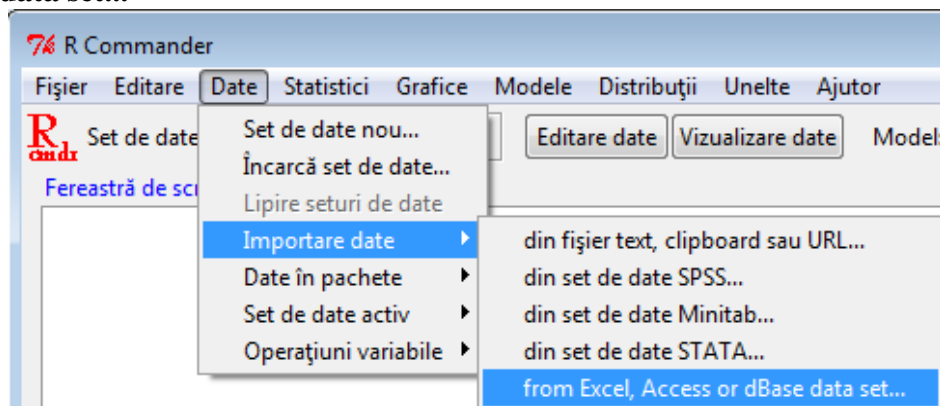
Urmați instrucțiunile de la adresa:

<http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/installation-notes.html>

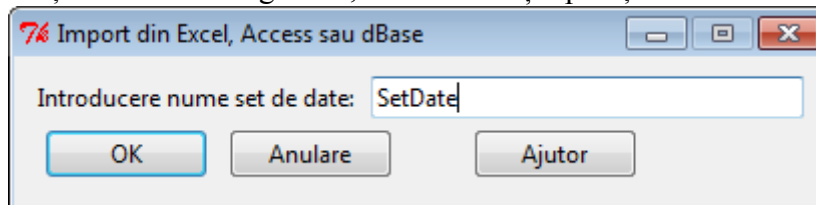
Importarea fișierelor Excel in R Commander

Înainte de a importa datele în Rcmdr, asigurați-vă că datele au fost corect culese și nu mai doriți să faceți nici o modificare asupra lor. De asemenea închideți fișierul de date înainte de a-l importa în Rcmdr!

Se merge în meniul **Date**, submeniul **Importare date**, opțiunea **from Excel, Access or dBase data set...**

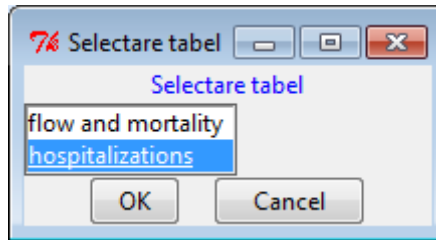


Apare o fereastră "Import din Excel Access sau dBase" în care suntem întrebați cum să denumim setul de date care va conține datele din fișier. În câmpul de editare existent în fereastră introduceți o denumire sugestivă, ex: SetDate și apăsați butonul OK.

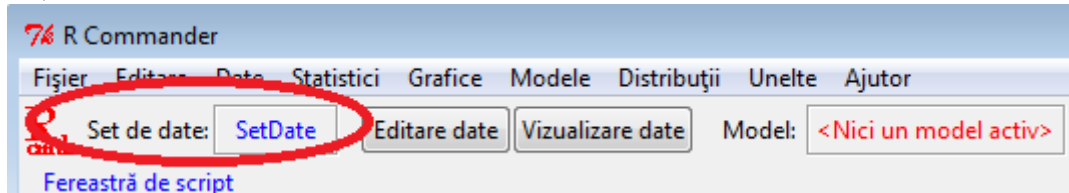


Apare o fereastră cu ajutorul căreia puteți să căutați fișierul cu date pe hard disk.

După ce l-ați localizat și l-ați deschis, dacă fișierul Excel conține mai multe foi de calcul, apare o fereastră care afișează ce foi de calcul există în fișier, și vă întreabă pe care din ele doriți să o importați. Selectați foaia care conține datele de interes și apăsați butonul OK.



În urma importării datelor în secțiunea Mesaje a ferestrei Rcmdr, vă scrie cum a decurs importarea (eventuale erori, modificări de denumiri de coloane, etc.). De asemenea în partea de sus a ferestrei Rcmdr se observă că setul de date activ se numește SetDate (sau cum l-am denumit).



Pentru a vizualiza datele existente în setul de date activ, se apasă butonul Vizualizare date, iar pentru editarea acestor date se apasă butonul Editare date, butoane aflate imediat sub meniul Rcmdr.

Modificarea variabilelor din setul de date

Schimbarea ordinii categoriilor unei variabile calitative

R commander ordonează alfabetic categoriile variabilelor calitative. Uneori dorim în grafice să avem categoriile într-o anumită ordine, sau în tabele. Dacă ordinea nu este corectă în tabelul de contingenta indicatorii vor oferi informații eronate. Dacă observați că softul statistic oferă indicatorii în alta ordine decât cea corectă/dorită – trebuie să o schimbați.

Astfel pentru un test diagnostic negativ este prima categorie și pozitiv a doua! Lucru incorrect în tabelul de contingenta. Pentru schimbarea ordinii se merge în meniul Date (Data) / Operațiuni variabile / Reordonare categoriilor... (Reorder factor levels). Alegeți variabila de interes. În câmpul denumirii variabilei (Nume factor (Factor name)) lăsați textul existent (în acest fel variabila va fi ordonată fără a-i se schimba numele). Bifați opțiunea creare factor ordonat (Make ordered factor). Apasați OK. O fereastră va întreba dacă doriți să înlocuiți variabila care există deja, apăsați Da. În noua fereastră apărută, în stânga sunt prezentate categoriile variabilei în ordinea lor inițială (Categoriile vechi/Old categories). În dreapta aveți câmpuri cu numerotarea acestora într-o ordine nouă (Noua ordine/New order). Pentru un test diagnostic în stânga ar fi negativ urmat de pozitiv. Atunci în dreapta ar trebui să modificăm să fie 2 și 1, pentru ca pozitiv să fie pe prima poziție. Apasăm OK și variabila este reordonată.

Descrierea numerică a variabilelor

Calitative

Vezi capitolul Realizarea tabelor subcapitolele Tabele de frecvență, Tabele de contingenta.

Cantitative

Datele cantitative se descriu în direct în text sau în tabele sub forma unor statistici simple (ex. medie, deviație standard, ...). În prealabil trebuie evaluată normalitatea datelor și

apoi se face descrierea. Normalitatea datelor se evalueaza cu ajutorul histogramei si a altor tehnici.

Se alege meniul Statistici (Statistics) / Sumarizari (Summaries) / Descriptive (Numerical summaries). In fereastra puteti selecta o variabila (sau mai multe variabile tinand apasata tasta CTRL). Puteti preciza ce statistici aveti nevoie (ex. pentru date normal distribuite bifati media si deviatia standard, pentru date non normal distribuite bifati quantile si lasati selectia (0, .25, .5, .75, 1). Pentru a obtine media si celelalte statistici pentru fiecare grup in parte dati click pe butonul Sumarizare pe grupuri (Group summary). In fereastra deschisa alegeti ca variabila de grupare (Group variable): Statina si apasati OK. Daca nu doriti statistici grupate pe grupuri nu folositi butonul de sumarizare pe grupuri. Apasati OK pentru a vedea statisticile selectate.

Teste statistice

Rezultatul testelor statistice este frecvent notat p-value dar si $\Pr(>F)$ sau variante similare.

Atentie uneori rezultatul e notat in format stiintific daca valorile sunt foarte mici. Ex. $4.6354e-9$. Acest rezultat indica $4.6354 * 10$ la puterea -9 , adica 4.6354 impartit la 10^9 , adica $0,0000000046353$. Daca vedeti p-value $< 2.2e-16$, inseamna ca p e foarte mic, mai mic decat $2.2 * 10^{-16}$.

Comparare date calitative / relatii intre date calitative

2/>2 esantioane independente

- (valori peste 5 în tabel așteptat în 80% din celule)
 - **Testul Hi pătrat (Chi square)**
 - Testul cu indicatori medicali (OR, RR, RA, Se, Sp, NNT,...): - vezi mai jos capitolul Calculul indicatorilor medicali: OR, ...
 - Testul simplu – fara indicatori medicali (OR, RR, RA):
 - În R Commander: meniu Statistici (Statistics)/ Tabele de contingenta (Contingency tables)/Tabel cu dublă intrare (Two way table)... În fereastra apărută se selectează la Testare ipoteze (Hypothesis tests) – Test Chi-pătrat de independenta (Chi square test). Variabila de grupare se ia din lista de variabile – pe randuri (rows – grouping variable), variabila dependenta se ia din lista de variabile – pe coloane (columns – dependent variable). Se selectează dacă se doreste sa se afiseze procentele pe linii sau coloane. Click pe OK.
- (valori sub 5 în tabel așteptat în peste 20% din celule –R (Rcmdr) ne sugerează dacă e necesar)
 - **Testul exact Fisher**
 - În R Commander: meniu Statistici (Statistics)/ Tabele de contingenta (Contingency tables)/Tabel cu dublă intrare (Two way table)... În fereastra apărută se selectează la Testare ipoteze (Hypothesis tests) – Test exact al lui Fisher (Fisher exact test). Variabila de grupare se ia din lista de variabile – pe randuri (rows – grouping variable), variabila dependenta se ia din lista de variabile – pe coloane (columns – dependent variable). Se selectează dacă se doreste sa se afiseze procentele pe linii sau coloane. Click pe OK.

2 esantioane dependente/perechi

- **Testul Mc Nemar**

- În R Commander se instalează plug-in-ul Rcmdrplugin.coin din meniul tools. În meniul Coin nou apărut se alege testul Marginal homogeneity test (pentru variabile dicotomiale oferind rezultatul testului Mc Nemar).

Comparare date cantitative

2 esantioane independente

- **Testul Student (t) pentru esantioane independente.** Are doua variante: cu variante egale sau cu variante inegale. Pentru a alege testul potrivit se poate folosi testul Levene pentru compararea variantelor. Se intra in meniul Statistici (Statistics) / Variante (Variances) / Testul Levene (Levene test...). In fereastra deschisa se alege la Factori (Factors), variabila de grupare (ex. variabila tratament), iar la Variabila Raspuns se alege variabila cantitativa de interes. Se apasa pe OK. Daca rezultatul testului $Pr(>F)$ – valoarea lui p e mai mica de 0,05 inseamna ca exista diferente intre variante – si se alege testul student pentru esantioane independente cu variante inegale, daca p e mai mare ca 0,05 se alege testul student pentru esantioane independente cu variante egale.
- Se intra in meniul Statistici (Statistics) / Medii (Means) / Testul t pentru esantioane independente (T test for independent samples...). In fereastra deschisa se alege la Factori (Factors), variabila de grupare (ex. variabila tratament), iar la Variabila Raspuns se alege variabila cantitativa de interes (ex. scadere colesterol). Se alege varianta de test t: cu variante egale sau nu selectant in zona: Asumare egalitate variante. Se apasa pe OK.

2 esantioane dependente/perechi

- **Testul Student (t) pentru esantioane dependente / perechi.**
- Se intra in meniul Statistici (Statistics) / Medii (Means) / Testul t pentru esantioane dependente (T test for paired samples...). In fereastra deschisa se alege la Prima variabila (First variable) variabila cu masuratorile pentru primul grup (sau prima masuratoare) (ex. colesterol initial), iar la A doua variabila (Second variabila) se alege variabila cu masuratorile pentru al doilea grup (sau a doua masuratoare) (ex. colesterol final). Se apasa pe OK.

Comparare date de supravietuire

- **Testul Logrank**
- (Pentru analiza de supravietuire plug-inul RcmdrPlugin.survival trebuie anterior instalat si incarcat – In varianta programului de pe site (vezi Instalarea simpla – recomandata !!!) – este deja instalat si incarcat). Pentru comparatii se utilizează meniul **Statistici**, submeniul **Survival Analysis**, comanda **Compare survival function...** În fereastra apărută în caz că nu au fost definite datele de supraviețuire trebuiesc selectate: variabila/ele care definesc timpul de supraviețuire (**Time or start/end times**), variabila care definește statusul cenzurii/evenimentele (**Event indicator**), și variabila de stratificare/grupare (**Strata**) – care va defini grupurile ce sunt comparate prin test. În caz că ele au fost definite anterior, ele sunt deja selectate în fereastră. În secțiunea **rho** se poate selecta tipul testului statistic care va compara grupurile din variabila de stratificare: 0 – testul log-rank (Mantel-Haenszel), respectiv 1 – testul Gehan-Wilcoxon modificat de Peto & Peto. **Atenție!** Pentru a nu avea erori la apăsarea butonului OK al ferestrei, dacă nu se selectează un subset al setului activ de date, atunci în secțiunea **Expresie de selectare** ștergeți complet conținutul câmpului de editare.

Calculul indicatorilor medicali: OR, RR, RA, Se, Sp, VPP, ARR, NNT, ...

Daca nu aveti meniul EBM: (Pentru calcula usor acesti indicatori plug-inul RcmdrPlugin.EBM trebuie anterior instalat si incarcat – In varianta programului de pe site (vezi Instalarea simpla – recomandata !!!) – este deja instalat si incarcat. Daca ati mers pe instalarea clasica, trebuie sa instalati pachetul RcmdrPlugin.EBM – vezi indicatiile de instalare pachete in capitolul Cum se instalează interfața grafică R Commander, ulterior plug-inul e pornit similar cu Pornirea interfeței grafice R Commander (Rcmdr)).

Pentru a calcula indicatori medicali utili in medicina, se acceseaza meniul EBM (Evidence Based Medicine). Din meniu se selecteaza tipul de studiu dorit (Diagnostic, terapeutic, prognostic). Fiecare astfel de meniu ofera indicatorii corespunzatori. In fereastra aparuta vi se ofera posibilitatea sa alegeti tipul de table de contingenta (procente pe linii, pe coloane), testele statistice dorite (Chi-patrat, Fisher,...). Esential trebuie selectata variabila de grupare (factorul de risc/prognostic la studiile de factori de risc; noul test diagnostic pentru studiile diagnostic; tratamentul pentru studiile terapeutice) si variabila dependenta (boala – rezultatul asteptat pentru studiile de factori de risc; testul standard de aur pentru studiile diagnostic; rezultatul asteptat pentru studiile terapeutice). Programul ofera un tabel cu numarul de subiecti si procentele pe linii sau coloane. Ofera testul/testele statistic/e ales/e si indicatorii medicali impreuna cu intervalele de incredere 95%. De asemenea va ofera si un grafic mozaic al relatiei dintre variabile.

Mare Atentie!!! La ordinea categoriilor pentru variabila de grupare si dependenta! Daca ordinea nu este corecta in tabelul de contingenta indicatorii vor oferi informatii eronate. Daca observati ca softul statistic ofera indicatorii in alta ordine decat cea corecta – trebuie sa o schimbati. R commander ordoneaza alfabetic categoriile variabilelor calitative.

Corelatii

Calculul coeficientilor de corelatie (Pearson/Spearman) si valoarea lui p a testului asociat lor

Coeficientul de corelatie (Pearson/Spearman) se pot obtine din R Commander, meniul Statistici (Statistics) / Sumarizari (Summary) / Matrice de corelatii (Correlation matrix). In fereastra aparuta se selecteaza variabilele intre care se doreste a se afla coeficientii de corelatie si p-ul asociat acestora. Trebuie selectate cel putin 2 variabile tinand tasta CTRL apasata. In cazul de fata se pot selecta toate cele 3 variabile. Daca datele sunt normal distribuite se alege coeficientul Pearson, altfel Spearman. (Normalitatea variabilelor poate fi evaluata cu ajutorul histogramelor si a altor tehnici) Pentru a obtine si p-ul asociat lor, se bifeaza optiunea: Valori p pairwise pentru corelatii Pearson si Spearman. Apasand OK se obtin 3 tabele. Primul tabel ofera o matrice cu coeficientii de corelatie – toate combinatiile posibile intre variabilele selectate (ex. Pentru 3 variabile selectate e un tabel 3 pe 3, diagonala fiind 1 – corelatia unei variabile cu ea insasi fiind evident 1). Al doilea tabel ofera p-ul asociat coeficientilor de corelatie din primul tabel. Daca e scris 0, asta inseamna ca valoarea lui p e foarte mica (scrieti $p < 0.001$ in loc de 0). Ultimul tabel este similar, dar p-urile oferite sunt ajustate pentru faptul ca s-au facut teste multiple. Luam valorile lui p din al doilea tabel. Dar e bine sa verificam daca exista diferente mari cu al 3-lea tabel. Prezenta acestora sugereaza ca intamplarea joaca un rol important in rezultatele obtinute si merita folosita ajustarea, respectiv p-urile din ultimul tabel.

Regresii lineare

Semnificația statistică a modelelor (p corespunzător lui **F significance**), coeficienții de determinare, precum și tabelele cu coeficienții variabilelor

Se obțin în R Commander din meniul Statistics / Fit models / Linear regression. În fereastra se alege variabila dependentă (ex. Greutatea) în casuta de selectare Response variable, și variabila/variabilele independente (ex. Lungimea) în casuta de selectare Explanatory variables. Dacă se dorește alegerea mai multor variabile independente, se ține apăsată tasta CTRL iar cu soarecul se aleg variabilele de interes. Se da OK. În zona de rezultate a ferestrei programului R Commander se observă un tabel al coeficienților regresiei. Coeficienții sunt trecuți în coloana Estimate. P-ul asociat coeficienților este trecut în coloana Pr ($>|t|$). Coeficientul de determinare r^2 și coeficientul de determinare ajustat este trecut după tabel (Multiple r-squared și adjusted r-squared) Pentru regresii simple (cu o singură variabilă independentă) se alege Multiple r-squared, dacă aveți regresii multiple alegeți adjusted r-squared. Semnificația statistică a modelului se află chiar la final p-value... Pentru a afla intervalul de încredere pentru coeficienții de regresie se alege din meniul Models / Confidence intervals. În fereastra aparută lăsați 95%. Veți obține un tabel atât cu coeficienții cât și intervalul de încredere 95% pentru fiecare din aceștia.

Regresii pentru date de supraviețuire

Regresia Cox – aflarea ratei hazardului

Dacă se dorește crearea unui model semiparametric Cox al hazardului proporțional se utilizează meniul **Statistici** submeniul **Adecvare (fit) modele** comanda **Cox-Regression Model...** Pentru analiza de supraviețuire plug-inul RcmdrPlugin.survival trebuie anterior instalat și încărcat – În varianta programului de pe site (vezi Instalarea simplă – recomandată !!!) – este deja instalat și încărcat.

În fereastra apărută în caz că nu au fost definite datele de supraviețuire trebuie selectate: variabila/ele care definesc timpul de supraviețuire (**Time or start/end times**), variabila care definește statusul cenzurii/evenimentele (**Event indicator**). Variabila de stratificare (**Strata**) – trebuie definită (una sau mai multe) dacă se dorește o analiză stratificată, utilă în cazul în care presupunția hazardului proporțional nu este îndeplinită pentru o variabilă. În caz că ele au fost definite anterior, ele sunt deja selectate în fereastră. Dacă era selectată deja o variabilă de stratificare, se poate deselecta apăsând tasta CTRL în timp ce se dă click pe variabila anterior selectată din câmpul de editare **Strata**. În secțiunea **Clusters** se pot selecta variabilele prin care se identifică clusterii pentru analiză, în caz că există așa ceva.

În secțiunea **Method for ties** trebuie precizată metoda care va fi utilizată pentru rezolvarea situațiilor cu date identice (dubluri). În mod standard se folosește metoda **Efron**.

În secțiunea **Robust standard errors** trebuie selectată metoda de calcul a erorii standard. Din oficiu se utilizează opțiunea care calculează erorile standard în mod robust (**Yes**).

În secțiunea **Variabile (dublu click către formulă)** se pot introduce termenii (variabilele independente) pentru modelul Cox. Prin dublu click pe variabilă aceasta este trecută în câmpul de editare al formulei modelului Cox, despărțite prin semnul plus.

În secțiunea **Expresie de selectare** se poate preciza o expresie prin care să se limiteze analiza pe un subset de observații din setul de date. Dacă nu se dorește acest lucru, trebuie șters textul din câmp: <toate cazurile valide>, altfel apare o eroare la rularea funcției.

În urma apăsării butonului **OK** obțineți descrierea modelului.

In regresia Cox intereseaza obtinerea HR (hazard rate – **rata hazardului**). In modelul de regresie acesta este notat $\exp(\text{coef})$. Valoarea lui p pentru fiecare coeficient este in coloana ($\Pr(>|z|)$), iar limitele intervalului de încredere 95% pentru rata hazardului (coeficienții la exponent) (lower .95 upper .95).

Realizarea tabelelor

R si R Commander ofera tabelele in format text pur, ceea ce face dificila copierea lor in documente de tip Word/Writer. De fapt pentru articolele stiintifice tabelele se fac separate in Word/Writer/Excel/Calc, preluand informatiile din alte tabele manual. Acest lucru este necesar intrucat datele din articole sunt sintetice, pe cand rezultatele programelor de analiza statistica nu sunt de obicei necesare integral.

Daca doriti sa creati un tabel in Word puteti sa utilizati tabul Insert – Table si sa alegeti un tabel de dimensiunile potrivite.

Puteti prelua tabelele totusi din Rcmdr in format html utilizand plug-inul RcmdrPlugin.Export – dar trebuie salvat rezultatul HTML intr-un fisier gol cu extensia .html. Acesta il puteti vizualiza intr-un browser si copia apoi in alt program.

Tabele de frecventa

Pentru a afla numarul si procentul pentru diverse categorii ale unei variabile folositi meniul Statistici (Statistics) / Sumarizari (Summary) / Distributii de frecvente (Frequency distributions). Alegeti variabila pentru care doriti aceste calcule si apasati OK. Programul ofera initial numarul de subiecti pe categorii (counts), si apoi procentele pe categorii (percentages). Datele se copiaza manual sau utilizand indicatiile de la inceputul capitolului Realizarea tabelelor.

Tabele de contingenta

Vezi capitolul anterior – Calculul indicatorilor medicali, precum si explicatiile de la inceputul capitolului Realizarea tabelelor.

Reprezentarea grafica a rezultatelor

Pentru a avea control mai bun asupra graficelor (posibilitatea precizarii titlului, denumirilor axelor, ...), trebuie utilizat un plug-in specializat in grafice, numit KMggplot2. Plug-in-ul se porneste accesand meniul Unelte (Tools)/ Inarca plug-in-uri Rcmdr (Load Rcmdr plugins)... In fereastra aparuta se selecteaza RcmdrPlugin.KMggplot2 si se apasa OK. Programul va va anunta ca plug-inurile vor fi disponibile numai dupa ce Commander va fi restartat. Apasati **Yes**.

Datele daca au fost deja importate nu au disparut. Daca dati click in bara de instrumente (situata sub bara de meniu), in dreptul Set de date (Data set), apare o fereastra de unde va puteti selecta baza de date importata.

Atentie graficele apar intr-o fereastra separata, dedicata graficelor.

Optiuni utile pentru grafice (marime text, culori, teme)

Plug-in-ul KMggplot2 ofera pentru graficele sale optiuni commune.

Daca doriti sa treceti **titlu**, puteti sa-l precizati in campul Title – dar pentru teze se recomanda ca titlul sa fie trecut sub grafic in forma text.

La **legenda** (Legend label) goliti casuta de text, sau treceti denumirea variabilei.

Puteti preciza **marimea caracterelor** (Font Size), tipul **font**-ului (Font pattern) si setul de culori preferat (Colour pattern) (aici puteti sa experimentati diverse variante din lista oferita).

Recomandam sa alegeti optiunea `theme_simple` – pentru grafice clasice, dar puteti incerca si alte teme din sectiunea Themes.

Optiunile Facet variables in rows/columns – sunt pentru reprezentari stratificate – le ignoram daca nu avem nevoie de ele.

Pentru variabile calitative / cantitative discrete

Descriere variabile cantitative

Grafice sectoriale (Pie)

Dupa ce ati pornit plug-inul KMggplot2 (vezi Reprezentarea grafica a rezultatelor), din meniul KMggplot2, se selecteaza Diagrama circulara (Pie chart). In lista Variabile (Variable) alegeti variabila pentru care doriti realizarea graficului. Urmatoarele 2 optiuni Facet variables in rows/columns – sunt pentru reprezentari stratificate – le ignoram daca nu avem nevoie de ele. Daca doriti sa treceti **titlu**, puteti sa-l precizati in campul Title – dar pentru teze se recomanda ca titlul sa fie trecut sub grafic in forma text. La **legenda** (Legend label) goliti casuta de text, sau treceti denumirea variabilei. Puteti preciza **marimea caracterelor** (Font Size), tipul **font-ului** (Font pattern) si setul de culori preferat (Colour pattern) (aici puteti sa experimentati diverse variante din lista oferita). Recomandam sa alegeti optiunea `theme_simple` – pentru grafice clasice, dar puteti incerca si alte teme din sectiunea Themes.

Graficul apare intr-o fereastră separata.

Evaluare legatura intre 2 variabile calitative

Dupa ce ati pornit plug-inul KMggplot2 (vezi Reprezentarea grafica a rezultatelor), din meniul KMggplot2, se selecteaza Bar chart for discrete variables (grafic coloane pentru variabile discrete). Esential trebuie selectata variabila de grupare - X variable (factorul de risc/prognostic la studiile de factori de risc; noul test diagnostic pentru studiile diagnostic; tratamentul pentru studiile terapeutice) si variabila dependenta - Stratum variable (boala – rezultatul asteptat pentru studiile de factori de risc; testul standard de aur pentru studiile diagnostic; rezultatul asteptat pentru studiile terapeutice). Se trec apoi denumirile axelor: Horizontal axis label (pentru variabila de grupare de pe axa x) si Vertical axis label (pentru a arata daca sunt procente sau numar de subiecti – vezi optiunea Axis scaling: percentages – pentru procente sau frequency counts – pentru numarul subiectilor/entitatilor). La Legend label – treceti denumirea variabilei dependente – selectata anterior in Stratum variable. In rest alegeti optiunile grafice care va ajuta, ex. din Themes – `theme_simple` – probabil mai placuta (vezi capitolul Optiuni utile pentru grafice).

Pentru variabile cantitative

Descriere date cantitative

Grafice tip histograma

Dupa ce ati pornit plug-inul KMggplot2 (vezi Reprezentarea grafica a rezultatelor), in meniul KMggplot2 se alege Histogram. In fereastră aparuta alegeti la Variable variabila pentru care doriti sa realizati histograma. La optiunea Axis Scaling alegeti Percentages, pentru a obtine histograma pentru valorile procentuale. In casuta de text Horizontal axis label – se trece denumirea variabilei cu unitatea de masura in paranteza, iar in Vertical axis label – se trece Procent (daca ati ales Percentages la Axis Scaling). Titlul nu se trece aici (doar in text sub grafic). Alegeti tema graficului (ex. `theme_simple`) si alte optiuni pentru forma graficului (vezi Optiuni utile pentru grafice).

Comparare date cantitative intre grupuri (in functie de o variabila de grupare)

Grafice de medii (error plot)

Dupa ce ati pornit plug-inul KMggplot2 (vezi Reprezentarea grafica a rezultatelor), din meniul KMggplot2, se selecteaza Box plot / Violin plot / Confidence interval (grafic coloane pentru variabile cantitative). Trebuie selectata variabila de grupare - X variable (ex. tratament), iar la X variable se alege variabila cantitativa de interes (ex. scaderea colesterolului). Se trec apoi denumirile axelor: Horizontal axis label (pentru variabila de grupare de pe axa x) si Vertical axis label (pentru variabila cantitativa). Atentie, nu uitati sa treceti unitatea de masura prescurtata in sistem international. La Graph type (tipul graficului) alegeti 95% CI (t distribution) – daca ati folosit testul Student pentru comparatia grupurilor – in situatia in care datele sunt normal distribuite. Daca doriti sa rotiti graficul alegeti optiunea Flipped coordinates. In rest alegeti optiunile grafice care va ajuta, ex. din Themes – theme_simple – probabil mai placuta (vezi capitolul Optiuni utile pentru grafice).

Grafice box plot (box and whiskers)

Dupa ce ati pornit plug-inul KMggplot2 (vezi Reprezentarea grafica a rezultatelor), din meniul KMggplot2, se selecteaza Box plot / Violin plot / Confidence interval (grafic coloane pentru variabile cantitative). Trebuie selectata variabila de grupare - X variable (ex. tratament), iar la X variable se alege variabila cantitativa de interes (ex. scaderea colesterolului). Se trec apoi denumirile axelor: Horizontal axis label (pentru variabila de grupare de pe axa x) si Vertical axis label (pentru variabila cantitativa). Atentie, nu uitati sa treceti unitatea de masura prescurtata in sistem international. La Graph type (tipul graficului) alegeti box plot – in situatia in care datele nu sunt normal distribuite. Daca doriti sa rotiti graficul alegeti optiunea Flipped coordinates. In rest alegeti optiunile grafice care va ajuta, ex. din Themes – theme_simple – probabil mai placuta (vezi capitolul Optiuni utile pentru grafice).

Evaluare legatura intre doua variabile cantitative

Grafice tip nor de puncte (scatter)

Dupa ce ati pornit plug-inul KMggplot2 (vezi Reprezentarea grafica a rezultatelor), in meniul KMggplot2 se alege Scatter Plot... In fereastra pe axa y (Y variable) se alege variabila dependenta (ex. Greutatea), iar pe x (X variable) se alege variabila independenta (ex. Lungimea). In casuta de text Horizontal axis label – se trece denumirea variabilei de pe axa x, iar in Vertical axis label – se trece denumirea variabilei de pe axa y. Titlul nu se trece aici (doar in text sub grafic). Se poate alege afisarea dreptei de regresie cu sau fara interval de incredere (Smoothing with/without C.I. (linear regression)). Alegeti tema graficului (ex. theme_simple) si alte optiuni pentru forma graficului (vezi Optiuni utile pentru grafice).

Pentru date de supravietuire

Descriere date supravietuire

Grafic Kaplan-Meier tabel probabilitate de supravietuire, mediana supravietuirii

Pentru a afisa un tabel de supravietuire (life table), sau pentru a crea un grafic de supravietuire se utilizează meniul **Statistici**, submeniul **Survival Analysis** (Analiza de supravietuire), comanda **Estimate survival function...** (Estimarea functiei de supravietuire)

În fereastra apărută în caz că nu au fost definite datele de supravietuire anterior trebuiesc selectate: variabila/ele care definesc timpul de supravietuire (**Time or start/end times**), variabila care definește statusul cenzurii/evenimentele – codat cu 1 pentru eveniment si cu 0 pentru cenzura (**Event indicator**), și variabila de stratificare/grupare (**Strata**).

La **Type of censoring** (tipul cenzurii) alegeți **Default**, dacă evenimentele apar la data înregistrării și nu într-un interval de timp anterior înregistrării.

La **Summary** selectați **Detailed** pentru a primi tabelul cu probabilitatea de supraviețuire.

În secțiunile legate de intervalele de încredere: **Confidence intervals** se recomandă opțiunea **Log**, în **Plot confidence intervals** (afișarea intervalelor de încredere pe grafic) se recomandă **Yes** (da) (default este no), iar în **Confidence level** (nivelul de încredere) se recomandă **95%**.

Dacă se dorește afișarea observațiilor cenzurate (recomandat), atunci se bifează opțiunea **Mark censoring times**.

Metoda recomandată pentru calculul estimării curbei/curbelor de supraviețuire este **Kaplan-Meier**, selectabilă în secțiunea **Method**.

Atenție! Pentru a nu avea erori la apăsarea butonului OK al ferestrei, dacă nu se selectează un subset al setului activ de date, atunci în secțiunea **Expresie de selectare** ștergeți complet conținutul câmpului de editare.

În urma apăsării butonului **OK** obțineți un grafic de supraviețuire Kaplan-Meier.

Tabel cu probabilitatea de supraviețuire

Pe lângă graficul de supraviețuire în fereastra de Output a R Commander, dacă ați selectat opțiunea Summary: Detailed, sunt listate mai multe informații: un tabel de supraviețuire (prezentat după comanda summary) cuprinzând timpul, numărul de subiecți la risc, numărul de evenimente, probabilitatea supraviețuirii, eroarea standard și limitele intervalului de încredere pentru probabilitatea supraviețuirii, pentru fiecare strat în parte al variabilei de stratificare/grupare.

Dacă doriți să aflați de exemplu **probabilitatea de supraviețuire la un an**, dacă timpul de supraviețuire este marcat în zile, atunci uitați-vă pe coloana time și identificați rândul imediat anterior sau egal valorii de 365 (un an). Pe acel rând puteți găsi probabilitatea de supraviețuire (**survival**) împreună cu intervalul de încredere 95% (limitele fiind: **lower 95% CI, upper 95% CI**).

Tabel cu mediana de supraviețuire

În fereastra de Output a R Commander, după comanda (> quantile), apar câteva tabele pentru fiecare strat/grup al variabilei de grupare. Mediana timpului de supraviețuire se află în coloana a doua (Survival probability - 0.5), rândul 2 (Estimate – quantile). Intervalul de încredere 95% corespunzător medianei se află în coloana a doua (Survival probability - 0.5), rândul 1 (Estimate – lower .95 CL), respectiv rândul 3 (Estimate – upper .95 CL).

Copierea / salvarea graficelor din R Commander in alte programe

Dupa ce ați realizat un grafic (vezi capitolele anterioare), îl puteți copia din fereastra care îl conține apăsând combinația de taste CTRL – c. Ulterior îl puteți lipi în Word/Writer sau alt program cu CTRL - v sau cu ajutorul soarecelui. Din fereastra rezervată graficelor din meniu puteți salva graficul în mai multe formate.

Puteti folosi și opțiunea Save Graph pentru graficele realizate cu KMggplot2. În acest fel după afișarea graficului, se deschide o fereastră care va permite să salvați fișierul în mai multe formate disponibile.

1) Fox J. R Commander Installation Notes [Internet]. Hamilton, Ontario:Mc Master University; c2010 [updated 2010 March 18; cited 2012] Available from: <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/installation-notes.html>