
Vág András
közgazdász, szociológus
E-mail: wifig@yahoo.com

Szabad hozzáférésű statisztikai elemző szoftverek a Világhálón

Az utóbbi években terjednek az Interneten a nyitott kódú szoftverek és ezzel párhuzamosan a szabadon hozzáférhető statisztikai elemző és modellező programok is. E mögött részben üzleti szándék rejlik, mivel az ingyenes vagy demoszoftverek a komolyabb, professzionális programok egyszerűsített vagy korlátozott funkciókkal rendelkező változatai. Szép számmal található a világhálón oktatási célokat szolgáló szoftverek, ezek mindegyike különböző egyetemek honlapjain szerepel. Egyre több, nemzeti kutatási alapokból finanszírozott fejlesztés eredményeként létrejött program érhető el az Interneten keresztül, mégpedig nemcsak az adott projekt tagjai számára, hanem világszerte bárkinek. Kapcsolódik ez a tendencia a nyitott kódú szoftverek terjedéséhez.

Valamelyest keveredik a „szoftver”, a „program”, az „alkalmazás”, a „fejlesztő környezet” és a „modell” kifejezés használata mind a felhasználók, mind a fejlesztők körében. A szoftver, a program és az alkalmazás általában egymás szinonimája, azokat a letölthető és installálható, vagy online futó eszközöket jelenti, amelyek valamilyen funkciókat látnak el vagy algoritmusokat futtatnak, de adatokat nem tartalmaznak, mint például a statisztikai elemző programok. A modellek kész struktúrákból vagy flexibilis modulokból álló reprezentációi valamilyen valós vagy képzelt jelenségnek, mint például a makroökonometriai modellek. A fejlesztő környezet olyan szoftver, amellyel modelleket lehet építeni, ilyenek például a mesterséges intelligenciát felhasználó eszközök. A szóhasználat azonban néha nem egyértelmű.

A programok jelentős része egyetemi portálokon található és alapvetően oktatási célból helyezték fel őket. Céges honlapokon is vannak statisztikai elemző programok, ezek elsősorban reklámcélokat szolgálnak. Az elemző programok sokféle célra felhasználhatók, de az előrejelzés-készítési tevékenységet egyik sem tudja maradéktalanul kiszolgálni. Indokolt és időszerű olyan portál kialakítása is, amely tekintettel van az előrejelzés-készítés speciális igényeire. Az Interneten egyébként elég sok, és szerencsére egyre több szabad hozzáférésű program található. Többségük nagyszerű-

en használható a különféle kurzusokban, egy-egy speciális módszer oktatásában és a kutatásban.

A szabad hozzáférésű statisztikai elemző programokkal kapcsolatos információkat három fejezetbe sűrítve mutatom be. Ezek a következők: a statisztikai elemzőkre mutató linkgyűjtemények, a letölthető „desktop” statisztikai adatelemzők és a webes – online futó – statisztikai adatelemzők.

Statisztikai elemző programok linkgyűjteményei

Ha nincs előzetes ismeretünk az online elérhető statisztika elemzőkről, vagy egyszerűen csak tájékozódni akarunk, akkor célszerű linkgyűjteményből választani, ahol könnyen áttekinthetjük a kínálatot.

Free Statistics (<http://freestatistics.altervista.org/>). A Free Statistical Software oldalain számos fontos hivatkozás található, ilyen például az adatelemző és kalkulátor linkgyűjtemény. Az oldalak megközelítőleg 600 hivatkozást tartalmaznak, ebből 380 online számításokat elvégző oldalakra mutat. A gyűjteményes oldalon számítástípusonként csoportosítva 10-20 hivatkozás található a leggyakrabban használt mintaösszehasonlító elemzésektől az összefüggés-vizsgálatokig. Az összeállításban a jövőkutatás szempontjából érdekes oldalak is találhatóak. Az említett linkgyűjteményen kívül érdemes átnézni a honlap többi részét is, mivel az rengeteg hasznos módszertani leírást, matematikai szoftvert és online adatelemzőre mutató hivatkozást is tartalmaz.

Guide (www.hmdc.harvard.edu/micah_altman/socsci.shtml). A hosszú nevű „The Impoverished Social Scientist's Guide to Free Statistical Software and Resources” (Az elszegényedett társadalomtudós kalauza az ingyenes statisztikai szoftverekhez és más forrásokhoz), *Micah Altman*, a Harvard egyetem egyik tanárának linkgyűjteménye, általános célú programcsomagokat, programkönyvtárakat és számos egyéb fontos hivatkozást tartalmaz.

Letölthető statisztikai elemzők

A következő alfejezet a jövőkutatásban hasznosítható, ingyenesen letölthető „desktop” statisztikai elemző programok neveit, internetcímeit és a szoftverek rövid ismertetését tartalmazza. Ezek mind olyan programok, amelyeket letöltés után installálni kell számítógépünkön.

Dataplot (www.itl.nist.gov/div898/software/dataplot/). A Dataplot, az Unix, a Linux és a Windows operációs rendszerek alatt futó tudományos vizualizációra, matematikai számításokra, statisztikai elemzésre és nemlineáris modellezésre használható szoftverrendszer. A programcsomag a következő – a jövőkutatásban is felhasz-

nálható – modulokkal rendelkeznek: adatsorok grafikai megjelenítése, elemzési célokat szolgáló grafikák, összefoglaló grafikák, alapstatisztikák, idősorelemzés, simítás, görbeillesztés, keresztábla-elemzés, valószínűség-számítás, többváltozós elemzések és matematikai függvények. A fejlesztők a szabadon letölthető programhoz további fejlesztési célokra a forráskódot is átadják.

Instat Plus (www.rdg.ac.uk/ssc/instat/instat.html). Az INTERactive STATistics programcsomagot általános statisztikai célokra fejlesztették, oktatási és kutatási célokra egyaránt alkalmas. A legutolsó, Windows operációs rendszer alatt futó (95, 98, NT, 2000) Instat Plus verziót kifejezetten a klímaváltozás kutatására bővítették (például Markov-láncok feldolgozását is tartalmazza).

IRRISTAT (<http://www.irri.org/science/software/irristat.asp>). Az IRRISTAT adatkezelésre és alapstatisztikai számítások elvégzésére alkalmas program. Fő alkalmazási területe a kísérleti eredmények kiértékelése. Emellett jövőkutatási célokra használható moduljai is vannak, például a korreláció és regresszió modul.

Mx (www.views.vcu.edu/mx/). Az Mx-program elsősorban mátrixalgebrai feladatok megoldására alkalmas. Több beépített strukturális egyenlet- és egyéb statisztikai modellező funkcióval rendelkezik. Ezen kívül illesztőfüggvényeket tartalmaz a LISREL-, LISCOMP-, EQS- és CALIS-módszerekhez. A felhasználók könnyen specifikálhatnak komplex „nem standard” modelleket, definiálhatják saját illesztőfüggvényeiket és optimalizálhatnak lineáris és nemlineáris egyenleteket.

R (www.r-project.org). Az R statisztikai számításokhoz és grafikákhoz fejlesztett programnyelv és programkörnyezet. Az R-program a Bell Laboratories által fejlesztett S- és S-Plus programnyelvhez hasonlít, ezért az R tulajdonképpen az S-programnyelv speciális implementációjának tekinthető. Van ugyan néhány különbség, de az S-programnyelven írott kódok többsége lefut az R-program alatt is. Az R-program széles körben kínál statisztikai (lineáris és nemlineáris modellezés, klasszikus statisztikai tesztek, idősorelemzés, klasszifikáció, klaszterezés stb.) és grafikai technikákat, emellett tovább is fejleszthető, mivel az R-program nyitott kódú. Az R-program egyik előnye, hogy könnyen tervezhető, publikálható ábrákat tud készíteni, beleértve a matematikai szimbólumokat és képleteket is, ha éppen ez szükséges. Jól használhatók a grafikai alapbeállítások, de ezeket a felhasználó minden szempontból megváltoztathatja. Az R-projekt hálózatának Magyarországon is működik tükörszervere a Semmelweis Orvostudományi Egyetemen (www.cran.hu).

Scilab (www.rocq.inria.fr/scilab/). A Scilab tudományos célokra fejlesztett programcsomag, elsősorban numerikus számításokhoz. Fontosabb algoritmusai és szolgáltatásai: adatstruktúrák feldolgozása (polinomok, mátrixok, többváltozós lineáris rendszerek); fejlett (a Matlab programhoz hasonló szintaxissal rendelkező) programnyelv; több száz beépített matematikai funkció; érdekes két- és háromdimenziós grafikai megoldások; szimuláció és optimalizáció; lineáris algebra; lineáris mátrix egyenlőtlenségek; beépített könyvtárak stb. A Scilab által nyújtott egyik érdekes

szolgáltatás a dinamikus rendszerek modellezését és szimulációját szolgáló „Scicos” interaktív környezet.

SSP (www.economics.pomona.edu/StatSite/SSP.html). A Smith's Statistical Package (Smith statisztikai csomagja) kisméretű, egyszerű, kifejezetten felhasználóbarát program. Az alapstatisztikákon kívül összefüggés-vizsgálatokat és grafikai megjelenítéseket is tartalmaz. Az adatok Excel-szerű táblákban szerkeszthetők.

Vista 5.5 (www.forrest.psych.unc.edu/research/). A Vista (Visual Statistics System – Vizuális statisztikai rendszer) jelmondata egyértelmű: „segít meglátni azt, amit az adatok mondanak”. A program a Linux-hoz hasonlóan nyitott kóddal rendelkezik, a fejlesztésbe be lehet kapcsolódni, Windows, Macintosh és Unix operációs rendszerek alatt egyaránt működik. A statisztikai vizualizációk dinamikusan és egy időben ábrázolják a különféle nézeteket. A vizualizációknak két fő típusa található a programban: a munkafolyamat lépéseinek megjelenítése (Work Maps) és az adatstruktúrában rejlő információk interaktív megjelenítése (Interactive Graphs). Ezen kívül – több más hasznos funkció mellett – adatbeviteli és szerkesztési modullal is rendelkezik.

Online statisztikai elemzők

Az online elemzők többsége Java programnyelven készült. Ezek a programok a böngészőben futnak, külön letöltésre nincs szükség, kivéve a legfejlettebb programokat, amikor szükség lehet a Java programnyelv legújabb verziójának installációjára. Az Interneten sok olyan honlap található, amelyre adatelemzőt tettek fel. Ezek többsége csak egyszerű számítások elvégzésére és az eredmények megjelenítésére alkalmas, leginkább a zsebalkulátorokhoz hasonlítanak. Egyes programok grafikákat készítenek, és már túllépnek a kalkulátor funkciókon, nem is beszélve a speciális (egyes esetekben mozgó) grafikákat készítő programokról.

Az online statisztikai elemzők nem tartalmazzak annyi funkciót és elemző algoritmust, mint a letölthető statisztikai programok. Ennek a viszonylagos egyoldalúságnak az az oka, hogy az adott programrészt minden használat alkalmával le kell tölteni, ami nagy programcsomag és rossz internet-hozzáférés esetén időigényes lenne. Egyes programozással azonban készíthető olyan program, amelyik kisméretű és mégis elég sok funkcióval és modullal rendelkezik. Az online statisztikai elemzők számára az adatokat – néhány kivételtől eltekintve – űrlapszerűen, mezőket kitöltve kell megadni. Minden számításához külön kell begépelni a számokat, ezért a felhasználásuk lassú és nehézkes. Részben az említett okból nagymennyiségű adat feldolgozását, vagy bonyolultabb számításokat ezekkel az eszközökkel nem lehet elvégezni. Valószínű, hogy ez a helyzet csak átmeneti és a jövőben egyre nagyobb és bonyolultabb feladatokat lehet majd online módon megoldani.

WebStat 3.0 (<http://webstatsoftware.com>). A Webstat 3.0 az egyik legjelentősebb online futó statisztikai elemző program. A program Java programnyelven készült. A Webstat 3.0 többféle adatkommunikációs menüponttal rendelkezik, többek között más Internet oldalakról is tud számításaihoz adatokat beolvasni. Ehhez természetesen az adatokat előre meghatározott formátumba kell rendezni. A program főbb statisztikai funkciói: összefoglaló statisztikák (sor-, oszlopstatisztikák, korreláció, kovariancia), táblázatok (gyakoriság, kontingencia), Z- és T-statisztikák, variancia-analízis és egyéb minta-összehasonlítások, regresszió (két- és többváltozós lineáris) és sokféle grafikus funkció is rendelkezésre áll.

Rweb (<http://www.math.montana.edu/Rweb/>). Az Rweb egy, az R statisztikai programnyelvhez kifejlesztett interfész, amely hasonlít az S- és S-plus programnyelvhez. Az Rweb modulok az R-programnyelv „mutat és klikkel” interfészei. Az adatkészlet leírását követően a programban két legördülő menü található: az első az elemzés típusának kiválasztására szolgál, a második pedig az adatkészlet kiválasztására. A portál tartalmaz saját, beépített adatokat, és a felhasználó a saját adatsorait is feltöltheti későbbi elemzés céljára. Az elemzési módszer és az adatok kiválasztása után hozzá lehet fogni az elemző oldal felépítéséhez. Az Rweb nyitóoldalán a program három különféle verzióját választhatja a felhasználó: 1. az alapverziót, amelyhez R-programozási ismeret szükséges; 2. a Javascript verziót, amihez a böngészőnek tudnia kell Javascript programot futtatni; és 3. az Rweb modulokat, amelyek „mutat-klikkel” interfésszel rendelkeznek és alapfokú statisztikai oktatásra fejlesztették ki.

Statiscope (www.df.lth.se/~mikaalb/statiscope/statiscope.shtml). A Statiscope leíró statisztikák számításához és az eredmények, illetve a grafikák megjelenítésére szolgáló Java nyelven készült program. A programba az adatok begépelhetők vagy egy Internetes címről letölthetők.

SISA (www.home.clara.net). A SISA- (Simple Interactive Statistical Analysis – Egyszerű interaktív statisztikai elemzés) egyszerűbb statisztikai számítások elvégzésére alkalmas, elsősorban minták összehasonlítására, eloszlások vizsgálatára, tehát közvetlenül nem használható a jövőkutató statisztikai kérdésfeltevésének megválaszolására.

Néhány, az Interneten található szabad hozzáférésű statisztikai szoftver elérhetősége

ADE-4	pbil.univ-lyon1.fr/ADE-4/
BIOMAPPER	www2.unil.ch/biomapper/
DATAPLOT	www.itl.nist.gov/div898/software/dataplot/
EPI DATA	www.epidata.dk/
FIRST BAYES	www.shef.ac.uk/~st1ao/1b.html

INSTAT PLUS	www.rdg.ac.uk/ssc/instat/instat.html
IRRISTAT	www.irri.org/irristat.htm
LISPSTAT	www.stat.uiowa.edu/~luke/xls/xlsinfo/xlsinfo.html
MIELKE AND BERRY STATISTICAL SOFTWARE	www.stat.colostate.edu/~mielke/permute.html
MIM31 STUDENT	www.hypergraph.dk
MX	views.vcu.edu/mx/
OPENSTAT VERSION 4	www.statpages.org/miller/openstat/
OX	www.nuff.ox.ac.uk/Users/Doornik/index.html
R	www.r-project.org
SCILAB	www-rocq.inria.fr/scilab/
SISA	home.clara.net
SSP	www.economics.pomona.edu/StatSite/SSP.html
STATCRUNCH	www.staterunch.com/
STATSCOPE	www.df.lth.se/~mikaelb/statiscopes/statiscopes.shtml
VISTA 5.5	forrest.psych.unc.edu/research/
WEBSTAT 3.0	webstatsoftware.com
WEKA 3	www.cs.waikato.ac.nz/~ml/weka/
WINBUG1.3	www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml
WINSAM	www.winsam.com/