

СЪВРЕМЕННИ ТЕНДЕНЦИИ В РАЗВИТИЕТО НА СТАТИСТИЧЕСКИЯ СОФТУЕР

Веселин Хаджиев¹

В статията се очертават основните тенденции в развитието на статистическия софтуер. За целта се използват признаците производител, хардуерна и софтуерна зависимост, интерфейс, степен на софтуерна осигуреност на статистическото изследване, специализация и пр. С цел да се улесни читателя е използвана и цитирана информация преди всичко от Интернет.

Разработването и предлагането на статистически софтуер през последните години нарасна толкова много, че броенето на новосъздадените и доразвити продукти практически загуби смисъл. Като основни фактори за бързия растеж могат да се посочат развитието на информационните технологии и Интернет, развитието на мащабите и формите на пазара със софтуерни продукти и не на последно място нарастването на търсенето и потреблението на статистически продукти.

Развитието и достъпността на информационните технологии като цяло създаде среда, в която стана възможна реализацията на доскоро теоретични методи и техники на изследване. Допълнителен тласък в разработването и предлагането на статистически софтуер даде развитието и достъпността на Интернет. Чрез глобалната мрежа стана възможна бързата и проста размяна на иновации и модули, чрез които се развиват и изграждат новите продукти. Развитието на формите на пазара, включително и електронните, премахна границите и доведе предлагането на софтуерни продукти практически до всеки. В крайна

сметка се създадоха условия за развитие и на самото търсене и потребление на статистически софтуер.

Предлагането на голям брой разнообразни по съдържание и функции продукти осигури възможност на потребителите да изберат и използват най-подходящи за целите на своето изследване софтуерни инструменти. От друга страна голямото разнообразие от статистически продукти създаде и трудности пред потребителския избор. Затрудненията произтичат от специфичния характер на статистическите софтуерни продукти, тяхната сложност и невъзможност да бъдат преценени предварително качествата им, необходимостта от значителна математическа и статистическа подготовка на потребителите и пр. Налице е нужда от очертаване на основните тенденции в развитието на статистическия софтуер и групиране на предлаганите статистически продукти по различни признаци. С цел да се улесни читателя при практическото използване на статията, е използвана и цитирана информация преди всичко от Интернет.

Каталози и класификации на статистически софтуерни продукти се публикуват в редица сайтове, например *STI Statistics*, *The Econometrics Journal Online*, *Q-Stats Software* и др. В повечето случаи предлаганите систематизации са по азбучен ред на имената на продуктите, имената на производителите и разпространителите, включени методи за анализ и в крайна сметка имат характер на каталози. Липсват систематизации по съществени признаци, въз основа на които да се очертаят съвременните тенденции в еволюцията на

¹ Д-р, главен асистент в катедра “Статистика” на Икономически университет – Варна, e-mail:

статистическия софтуер. В тази статия ще се разгледа развитието на статистическите софтуерни продукти по признаците *производител, хардуерна и софтуерна зависимост, интерфейс, степен на софтуерна осигуреност на статистическото изследване, специализация* и пр.

Първата тенденция, която се очертава веднага, е непрекъснатото разширяване на кръга от производители на статистически софтуер. Тази констатация може да се обясни както с глобализацията и достъпността на информационните ресурси и инструменти, така и с нарастването на специфичните нужди от статистически софтуер. Като се има предвид, че различните производители си поставят различни цели при разработването на даден продукт, се заключава че и техните продукти ще притежават различни свойства. Най-общо производителите могат да се разделят на следните групи: университети, научни институти, комерсиални производители и др.

От направеното проучване се установи, че много голям брой софтуерни статистически продукти се разработват от университетите. Тази констатация се обяснява със следните предпоставки и условия. Първо, университетите притежават необходимия научен потенциал за разработка както на универсални, така и на специализирани софтуерни продукти. Второ, самите университети са голям потребител на статистически софтуерни продукти. Това се свързва както с извършваните в университетите изследвания, така и с провеждания образователен процес. Характерно за университетските продукти е това, че те не са анонимни и имат водещи автори – математици, статистици и специалисти в дадена предметна област. Към по-

известните университетски статистически софтуерни продукти могат да се споменат *Ox* и *PcGive* на университета в Оксфорд (<http://hicks.nuff.ox.ac.uk/Users/Doornik>), *BUGS* на университета в Кембридж (<http://www.mrc-bsu.cam.ac.uk/bugs>), *COUNT*, *EI*, *EzI*, *Maxlik* на Харвардския университет (<http://gking.harvard.edu>), *SHAZAM* на университета в Британска Колумбия (<http://shazam.econ.ubc.ca>), *STAT* на университета в Сан Диего, Калифорния (<http://www.cis.ohio-state.edu/~perlman/statinfo.html>) и др.

Друг голям производител на статистически софтуерни продукти са изследователските институти. Тази констатация се обяснява както с наличието на значителен научен потенциал в тези институти така и с необходимостта от разработването на специфични софтуерни продукти, използващи специфични по свойства информация и специфични статистически методи за анализ и прогнозиране. Към известните статистически софтуерни продукти на изследователски институти могат да се причислят *CrossGraphs* на Belmont Research Inc. (<http://www.belmont.com>), *DOE KISS*, *SPC KISS* на Air Academy Associates (<http://www.airacad.com>), *CPS Utilities* на Unicon Research Corporation (<http://www.unicon.com>), *NOSTRADAMAS* на New Energy Associates (<http://www.newenergyassoc.com/html/nost.html>), *Prophet* на National Center for Research Resources (<http://www.prophet.abtech.com>), *SCA* на Scientific Computing Associates Corp. (<http://www.scausa.com>) и др.

Третата голяма група създатели на статистически софтуер са комерсиалните производители. Тук трябва да се направи уточнението, че почти всички производители предлагат своите продукти и на пазара. Но докато университети, изследователски институти и пр. използват създадените продукти и за други цели, комерсиалните производители

създават своите продукти единствено с цел продажба. Характерна черта на комерсиалните статистически продукти е, че техните създатели, независимо от предметната ориентираност, се стремят да покрият по възможност по широк кръг от статистически методи. Това е израз на стремежа да бъдат привлечени по-широк кръг от клиенти с разностранни потребности и интереси. Към по-известните статистически софтуерни продукти могат да се причислят **EXPO** на Leading Market Technologies (<http://www.lmt-expo.com/>), **Forecast Pro** на Business Forecast Systems (<http://www.forecastpro.com/>), **LIMDEP** на Econometric Software (<http://www.limdep.com/>), **SAS**, **JMP**, **StatView** на SAS Institute (<http://www.sas.com/>), **SPSS**, **SYSTAT**, **BMDP**, **SigmaStat** на SPSS Inc. (<http://www.spss.com/>), **Statistica** на StatSoft (<http://www.statsoft.com/>) и много други.

Четвърта обособена група от производители е групата на индивидуалните автори. По правило създателите на софтуерни продукти са водещи преподаватели и учени от университети и институти. Разбира се отделянето на тази група е сравнително условно, тъй като авторите използват и ресурсите на съответните институции. Характерно за тези продукти е, че те са по-тясно специализирани за дадена група методи или дадена предметна област, в която са специализирани и самите автори. Към по-известните продукти могат да се причислят **Microfit** на Prof. M.H. Pesaran и Dr. Bahram Pesaran (<http://www.intecc.co.uk/camfit/>), **MLn** на Prof. Harvey Goldstein (<http://www.ioe.ac.uk/multilevel/>), **TDA** на Prof. Dr. Goetz Rohwer и Dr. Ulrich Poetter (<http://www.stat.ruhr-uni-bochum.de/tda.html>), **Applied Stat** на David Cooper и Associates (<http://members.aol.com/djpcoper/spc.htm>) и др.

Съществена тенденция в развитието на статистическия софтуер е нарастващата независимост от операционната система и хардуера. Почти всички производители предлагат своя софтуер както за различни класове и типове компютри, така и за различни операционни системи. По този начин се осигурява достъп на повече потенциални клиенти до продуктите на производителя. Обичайна практика е да се предлагат едни и същи продукти във версии за PC и Macintosh, например продуктите на ACL Software (<http://www.acl.com/>), Abacus Concepts Inc (<http://www.abacus.com/>) и др. Още по-голямо разпространение е намерило предлагането на версии на продукти за различни операционни системи – Windows, DOS, UNIX и пр. Като пример могат да се посочат продуктите *BACC 1999* (<http://www.econ.umn.edu/~bacc>), *TSP* (www.tspintl.com) и др. Във връзка с констатираната тенденция се наблюдава и сближаване на потребителските интерфейси на продуктите под различни операционни системи. Следваща стъпка в премахването на различията между операционните системи и хардуер е прилагането на протоколите на глобалната мрежа и използването на статистически софтуер директно в средата на Интернет, например *SHAZAM*. Следователно от гледна точка на потребителя факторите хардуер и операционна система губят своето решаващо значение при избора на статистически софтуер.

В развитието на статистическия софтуер се констатира отчетлива тенденция на сближаване на интерфейса на продуктите от типа “графичен” и типа “командноориентиран”. Това се изразява в интегриране на предимствата на двата подхода при изграждане на комуникациите с потребителя. От една страна чрез възможностите на

менютата се изгражда интуитивен и лесноразбираем интерфейс, характерен за повечето приложения под Windows. Обикновено по този начин се реализира използването на класическите статистически методи – дескриптивна статистика, дисперсионен анализ, регресионен анализ и пр. Така създадените продукти се надграждат с модули от програми на подходящ команден език. Чрез тях се осигурява гъвкавост на софтуерния продукт, бърза реализация на нови статистически методи, осигуряване на интерактивност на обработката на данни, възможност за използване на on-line библиотеки с програмни модули и пр. Като пример за интегриране на двата подхода при изграждането на интерфейса могат да се посочат продуктите на Stat Soft. В тях наред с традиционната комуникация от типа “графичен” се включват модули на командните езици “BASIC” и “STATISTICA COMMAND LANGUAGE”. Например за надграждане на базовата конфигурация от модули производителят предлага програми за извикване на изпълними файлове в средата на статистическия продукт, оценки на Vox-Cox трансформации и др.

Трябва да се отбележи, че при командноориентирания интерфейс е необходимо много високо ниво на специална подготовка. Чрез него се организира манипулацията на данни и пресмятанията на характеристики на най-ниско ниво. При интерфейса от типа “графичен” се изисква по-общо ниво на подготовка. Чрез него се реализират отделните етапи, методи и етапи от методи, но организацията на цялото изследване се извършва от потребителя.. При анализирането на използваните от статистическите продукти интерфейси трябва да се отбележи и появяването на нов тип

комуникации с потребителя. Става въпрос за продукти с графичен интерфейс с много опростена комуникация и неизискващи специална статистическа подготовка от потребителя. Към тази група софтуер може да се причисли *ViSta* (<http://forrest.psych.unc.edu>).

Съществена тенденция в развитието на статистическите продукти е стремежа на част от производителите да осигурят по-висока степен на софтуерна обезпеченост на статистическото изследване. Това се изразява в разработването на специализирани продукти за отделни етапи или универсални продукти за цялото статистическо изследване. В зависимост от това, каква част от статистическото изследване се осигурява от съответните статистически софтуерени продукти, се оформят следните групи:

- статистически софтуерни продукти, чието основно предназначение е да осигурят събирането и подготовката за последваща обработка на статистическите данни;
- статистически софтуерни продукти, чието основно предназначение е да осигурят обработката на статистическите данни чрез подходящи статистически методи;
- статистически софтуерни продукти, чието предназначение е да осигурят цялото статистическо изследване;
- софтуерни продукти за конвертиране и обмен на информация между бази данни, електронни таблици и статистически софтуерни продукти.

Към първата група се причислява софтуерът за създаване и поддръжане на бази от статистически данни. Освен универсалните за

тази цел софтуерни продукти като *dBASE*, *Paradox* и др. се разработва специализиран софтуер, чрез който се създават и подържат данни, притежаващи свойства на статистическа информация. Характерно за този тип софтуер е, че чрез него се извършва първичното събиране на данни от обекта на изследване. Например наблюдение на реализираните продажби чрез баркодове, наблюдение на заявки, продажби и други обекти, извършвани чрез Интернет, наблюдение на транзакции извършвани по електронен път (например на стоковите и фондовите борси), непосредствено измерими чрез подходящи инструменти признаци (например хидроложки, метеорологични и прочие данни), непосредствено измерими данни в индустрията (например системите за контрол на качеството) и пр. Към тази група спадат продуктите *SurfReport* (<http://www.bienlogic.com/SurfReport>), *WatchWise* (www.watchwise.com), *Pangsway Limited* (www.pangsway.demon.co.uk/pangsway.htm) и др.

Към втората, най-многобройна група спадат продуктите, чието основно предназначение е да осигуряват обработката на статистическите данни чрез подходящи методи. Трябва да се отбележи, че в тази група попадат и първообразите на статистически софтуерни продукти. Характерна особеност на тези продукти е, че се използват данни от вторичното наблюдение. Основният акцент е поставен върху методите за статистически анализ и прогнозиране. Към тази група спадат продуктите на SPSS, StatSoft, Stata Corporation (www.stata.com) и пр.

Акцентите в третата група от статистически продукти е поставен върху цялото изследването. Характерно за тази група е, че чрез него се осигурява целия процес на събиране на първични данни, обработка,

интерпретация и препоръки. Тези системи имат характер на съветващи on-line системи и съдържат елементи на изкуствен интелект. Трябва да се отбележи, че на тези продукти липсва характерната за предната група гъвкавост. Към тях могат да се причислят продуктите на Raosoft (www.survey-software.com), STEN Information Resources (www.steninfores.com) и др.

Статистически данни се събират и организират чрез най-различни софтуерни продукти, представени чрез различни формати. Това затруднява обмена на данни, оскъпява продажбата на информационни продукти и пр. За решаване на този проблем и улесняване обмяната на данни се разработват специални конвертиращи и транслиращи софтуерни продукти. Те подържат много широк спектър от формати и маски на различни бази от данни, електронни таблици и статистически софтуерни продукти. Към тази четвърта група се причисляват продуктите *Circle Systems* (www.stattransfer.com), *Conceptual Software* (www.conceptual.com) и др.

Крайна цел от приложението на всеки статистически софтуер е да се създадат пряко или косвено условия за прогнозиране на бъдещо състояние и развитие на изследвания обект. От тази гледна точка могат условно да се очертаят три тенденции в развитието на статистическия софтуер: описателни системи, аналитични системи и автоматични системи за прогнозиране.

Описателните системи осигуряват по-ниска степен на обработка на информацията, включват предимно графични методи, описателна статистика и пр. Чрез тези продукти, например модула “*Analysis ToolPak*“ на MS Excel се постига предимно описание на текущото състояние на обекта чрез подходящи статистически характеристики.

Чрез аналитичните системи се извършва задълбочено статистическо изследване, оценяват се различни статични и динамични характеристики въз основа на които потребителят задава параметри и произвежда прогнози за бъдещо състояние на обекта. Като пример могат да се посочат продуктите на Quantitative Micro Software *Eviews* (<http://www.eviews.com>). Автоматичните системи за прогнозиране дават като резултат готови оценки за бъдещо състояние на обекта. За потребителите остават скрити всички междинни етапи на статистическото изследване. Автоматичните системи за прогнозиране не са гъвкави като аналитичните системи. Те могат да се прилагат само в случаите, когато условията за развитие на обекта са постоянни и вариацията по прогнозираните признаци е относително хомогенна. Като пример за автоматична система може да се посочи продукта *Autobox* на Automatic Forecasting Systems (<http://www.autobox.com>).

При разработването на статистически софтуерни продукти се констатира противоречиви тенденции относно предметната област. От една страна продължава разработването на все по-специализирани продукти. Например в областта на хидроложките изследвания се предлагат продуктите *Amos* на SmallWaters Corporation (<http://www.smallwaters.com/amos>), *Aqua Chem* на Waterloo Hydrogeologic Inc (<http://www.flowpath.com>), в областта на медицинските изследвания се предлагат продуктите *EpiCalc 2000* на Brixton Health (<http://www.myatt.demon.co.uk>), *Epi Info* на Centers for Disease Control and Prevention (<http://www.cdc.gov/epo/epi/epiinfo.htm>), в областта на спортната статистика се предлагат продуктите *Team Sport Software* (<http://www.teamsportsw.com/>), *Sport Shareware* (www.absnet.com) и др.

При класифицирането на един продукт за специализиран в дадена област се имат предвид преди всичко включените методи за статистически анализ. Тази практика не е съвсем удачна, тъй като е известно че статистическите методи са инвариантни по отношение на предметната област. Ако дадена статистическа информация притежава необходимите свойства да бъде обработена чрез даден статистически метод, то това може да се осъществи както със *Statistica*, *SPSS*, *Eviews*, така и с *BUGS*, *NOSTRADAMAS* и пр. От тази гледна точка групирането по предметна област само според включените статистически методи за анализ не е достатъчно. Разбира се не може да се отрече, че статистическите данни за икономиката притежават предимно едни свойства, статистическите данни в биологията и медицината притежават предимно други свойства, статистическите данни в психологията притежават трети свойства и пр. Същественният критерий, според който продуктите трябва да се класифицират в различни предметни области е подготовката на данните и извличането от тях на едни или други свойства. От тази гледна точка строго специализирани продукти, например в областта на статистическия контрол на качеството, са продуктите на Hertzler Systems Inc (www.hertzler.com), продуктите на IMPACT (<http://www.iaehv.nl/users/mschaeff>) и др.

От друга страна се развива и тенденцията да се разработват и предлагат мощни универсални продукти. Към тях могат да се причислят споменатите по-горе *Statistica*, *STATGRAPHICS* и др. Тъй като включването на много широк кръг от методи и техники прави продуктите тежки за използване и неприемливо скъпи, се налага да се прави частична специализация. Производителите решават този

проблем принципно по два начина. Единият начин е създаването на една обща платформа, осигуряваща манипулация на данните и много на брой специализирани модули. Пример за използването на такъв подход са продуктите на Stat Soft. Другият начин е създаването на общи конвенции за работа с данните, еднотипни модули за извършване на изчисленията и фамилия от статистически софтуерни продукти с различна ориентация. Пример за такъв подход са продуктите на SPSS Inc. които наред с основния продукт *SPSS* създават и съпътстващи продукти като *SYSTAT*, *SigmaStat*, *SigmaPlot*, *AutoSignal* и пр.

Интензивното и многопосочно развитие на статистическия софтуер създава предпоставки за решаване на все по-широк клас от изследователски задачи. От своя страна голямото разнообразие от продукти затруднява избора на потенциалните потребители и възниква необходимост от подходяща систематизация. Очертаните по-горе посоки в развитието на статистическия софтуер са отправни точки в изграждането на класификации по съществени, интересувачи най-вече крайните потребители признаци и провеждане на сравнителни изследвания за еднотипни продукти. По този начин ще се осигурят необходимите условия за правилен подбор и използване на статистически софтуерени инструменти.

Използвани източници: Гоев, В. Какво трябва да знаем за статистическите пакети приложни програми? Сп. Статистика, С., 1997 г., бр. 1, с. 40-45;
<http://www.eur.nl>, <http://www.stats.gla.ac.uk/>, <http://www.stat.ufl.edu/>, <http://www.ai-geostats.org/>,
<http://www.cpcug.org>, <http://tolstoy.nsw.cmis.csiro.au/>.

СЪВРЕМЕННИ ТЕНДЕНЦИИ В РАЗВИТИЕТО НА СТАТИСТИЧЕСКИЯ СОФТУЕР

Веселин Хаджиев²

Резюме

Разработването и предлагането на статистически софтуер се разви изключително много през последните няколко години. За това способстват редица фактори, като развитието на информационните технологии и Интернет, развитието на мащабите и формите на пазара, нарастването на потреблението на статистически продукти и пр. Голямото разнообразие от продукти с различно съдържание и функции от своя страна се явява като проблем пред избора на потребителя. Обосновава се необходимостта от очертаване на основните тенденции в развитието на статистическия софтуер. С цел да се улесни читателя при практическото използване на статията, е използвана и цитирана информация преди всичко от Интернет.

В изложението по-нататък се разглежда развитието на статистическия софтуер по групи производители, по хардуер и операционна система, по интерфейс, по степен на осигуряване на статистическото изследване със статистически софтуер, по специализация и пр. Обосновава се необходимостта от разработване на класификация на статистическите продукти по съществени признаци и провеждане на сравнителни изследвания.

² Д-р, главен асистент в катедра “Статистика” на Икономически университет – Варна, e-mail: vesel6@dir.bg