JOURNAL OF APPLIED
COMPUTER SCIENCE
Vol. 19 No. 1 (2011), pp. 65-83

Free and Open Software for Applied Statistics. A Comparison and a Case Study in Biometrics

Ewa Niewiadomska[1], Adam Niewiadomski[2]

[1]*Medical University of Silesia, Department of Biostatistics*
Piekarska 18, 41-902, Bytom, Poland
e.j.niewiadomska @ gmail.com

[2]*Institute of Information Technology*
Technical University of Łódź
Wólczańska 215, 90-924, Łódź, Poland
Adam.Niewiadomski @ p.lodz.pl

Abstract. *This article is intended to be a position paper on advantages of free and open software for statistics and its applications to biometrics and biostatistics. Especially, the authors focus on the R package viewed as a new and still insufficiently recognized or received by the scientists, researchers, students, etc. Sample statistical computations and tests in biometrics are presented, and the most common functions and procedures are analysed and compared. Although this is a position paper from the point of view of applied computer science, the authors briefly present some original results within applied statistics (in particular: biometrics) and related computational methods using dedicated software.*

Keywords: *free software, open software, GNU/GPL software, applied statistics, statistical computations, R package, R language, literate programming, biometrics, biostatistics.*

1. Introduction

For last 80 years, at least, statistical computations are one of the most precise, commonly used and approved methods for collection, organization, analysis, and interpretation of data. The so-called *applied statistics* is a science that contracts statistical methods and other domains, e.g. biometrics, demography, econometrics, or statistical physics. Obviously, these methods must be supported by information technology, computation techniques and dedicated software packages. Unfortunately, many software products for applied statistics is hardly accessible, mostly because of high costs, e.g. MatLab [1], SPSS [2], SAS [3], Minitab [4], and other. Besides, most of them are not open software, which means that plugins or additional applications and features can be created only by producers, but not by communities of users or just enthusiasts.

Hence, the scope of this paper is to present information on still more and more popular **free and open** software packages for statistics and on using these packages in practice. Especially, the authors concentrate on the R package and language [5], mostly because of its scalability, programmability, and many features e.g. numerous plugins and possible conversion of output to LaTeX.

The paper is organized as follows: Section 2 collects basic information on commonly used software packages for applied statistics, which is intended to give a background for the presentation of the R package in Section 3 which also presents a comparison of features and options of presented software packages. A case study: using the R package and R programming language in sample statistical computations in biometrics, is presented in Section 4. Then, the paper is concluded in Section 5.

2. Free and open software for applied statistics

2.1. Gnumeric

Gnumeric is a free software (that runs under the Gnome Project) for statistical spreadsheets [6]. Especially, the so-called *Data Analysis Module* is expanded and complex. It provides 500 statistical procedures and functions, similar to those known from Visual Basic in e.g. MS Excel 2000. Statistical functions and tests (the Statistical Analysis package) are based on those known from the R language (see Section 3).

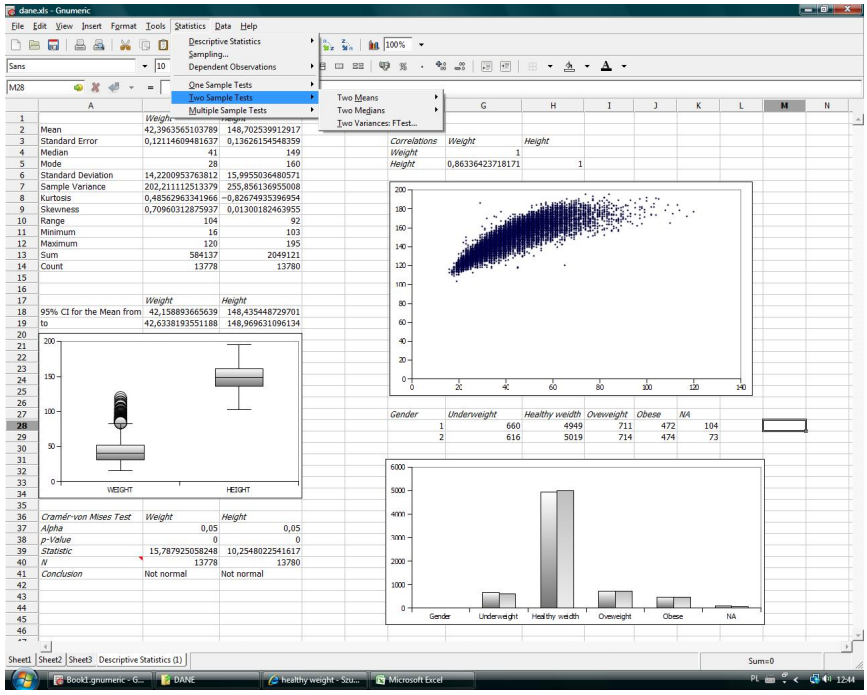

Figure 1. A screenshot of Gnumeric

Gnumeric reads files saved as spreadsheets in the following formats: MS Excel, OpenOffice Calc, Lotus 1-2-3, Applix, Psion, Sylk, XBase, Oleo, PlanPerfect, Quattro Pro, and as HTML. Gnumeric saves spreadsheets as: `.xls` for MS Excel, `.ods` (i.e. Open Document Spreadsheet) for OpenOffice Calc, as the LaTeX source and as HTML. The most recent version of Gnumeric is v.1.10.15 (June 20, 2011, see [6]).

User interface, formal language, and licences The user interface is organized similarly to a typical spreadsheet (see Figure 1). It contains 2^8 columns and 2^{16} rows. Gnumeric is accessible in 50 language versions, including Polish. A complete description of functions in Gnumeric is accessible in The Gnumeric User's Manual [6].

Gnumeric is free, accessible in terms of GNU General Public Licence (GPL), see: [6]. Updates of Gnumeric may be found on [7].

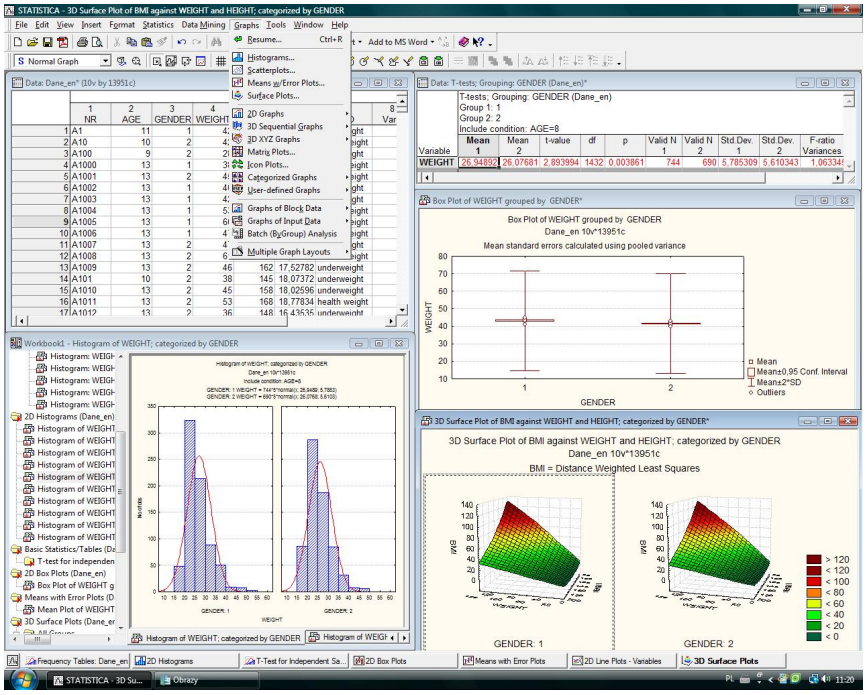

Figure 2. A screenshot of Statistica trial version

2.2. Statistica v. 8.0

Statistica is a system produced by StatSoft [8]. The software provides more than 10 000 functions for statistical computations, especially for data analysis, statistical procedures and algorithms, and methods of graphical representation of results. Although it is a commercial and not open software, we present it mostly because of its popularity.

The newest version is Statistica v.9.0 (June 20, 2011). Statistica enables opening spreadsheets saved in the following formats: MS Excel, SPSS, SAS, Quattro Pro, HTML. Statistica writes spreadsheets as: .sta – Statistica, `.xls` – MS Excel, `.sav` – SPSS, `.sas7bdat` – SAS, and as HTML.

User interface, formal language, and licences The Statistica system can be explored in three ways. They provide different user interfaces (it is worth noticing

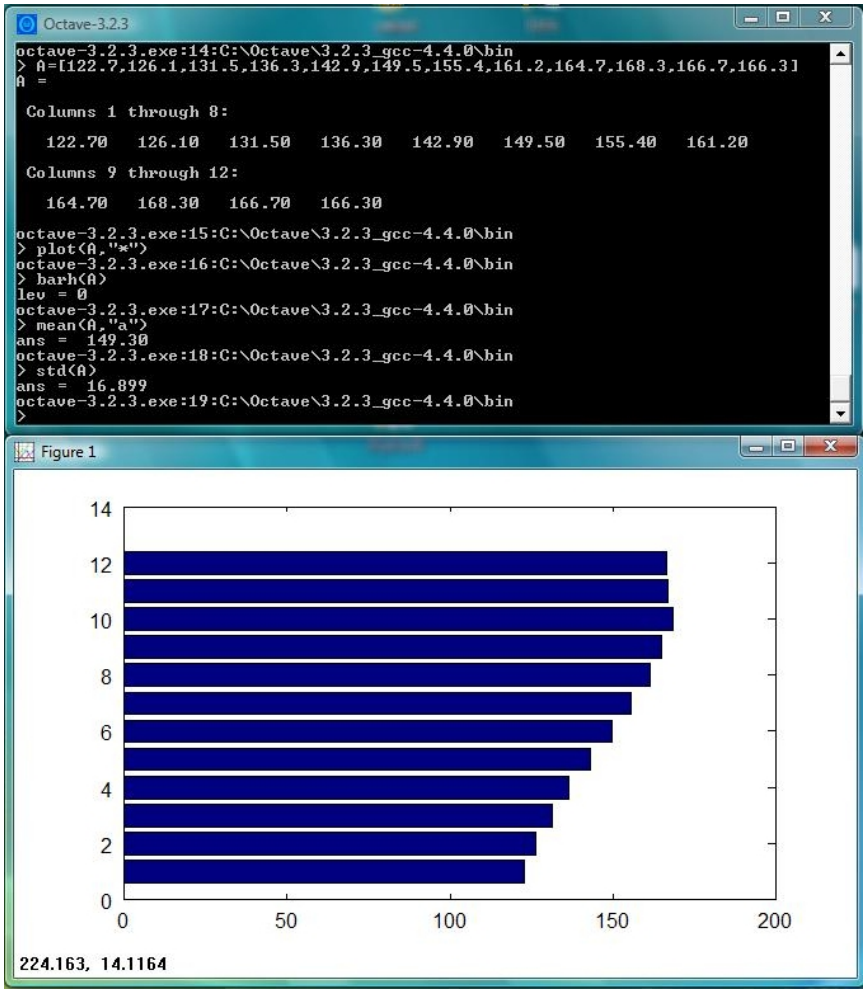

Figure 3. A screenshot of GNU Octave (the upper part), and a sample chart generated as a graphical output (the lower part)

that they are all Graphical User Interfaces – GUI):

1. The interactive mode (also called *conversational*) – in this mode the interface is similar to that known from spreadsheets (MS Excel, OpenOffice Calc), see Figure 2;

2. The GUI based on *STATISTICA* Visual Basic programming language;
3. The GUI mode based on a Web browser.

Statistica uses its own programming language *STATISTICA* Visual Basic, which is similar to the commonly known version of Visual Basic. Macros – sequences of commands – and other elements of spreadsheets can be developed using the C++, Java, or R (from Statistica 9.0) programming languages. The interfaces are presented in Figure 2. The only free version of Statictica 9.0 is 30 day trial software.

2.3. GNU Octave

GNU Octave is an environment for numerical computations. It is created by John W. Eaton in 1996, inspired by some MatLab functions for applied statistics, cf. [1, 9].

User interface and formal language Commands of Octave are entered in the console mode, and executed immediately (*interactive mode*). Octave does not provide any graphical user interface. Scripts – sequences of commands – are created with a text editor in a console mode, see Figure 3. The lower part of Figure 3 illustrates a sample chart: a graphical output generated by GNU Octave.

This software is accessible only in English language version. GNU Octave is freely accessible in terms of General Public License (GNU GPL) [10]. The manual is published by Free Software Foundation, Boston as a handbook [9].

2.4. Other packages

Because of the limited space of this paper, the authors are forced to omit many other free and/or open software packages for applied statistics, e.g. OpenEpi [11] or PSPP (free replacement for SPSS) [12]. A list of over 80 free software packages for applied statistics is accessible at `http://statpages.org/javasta2.html` [13].

3. R

The R package, or briefly: R is an environment for statistical computations created by Ross Ihaka and Robert Gentelman. The first version of R was presented in 2000. Currently, it is developed by the international "R Core Team". Installation versions and additional packages are accessible on the website of the R project

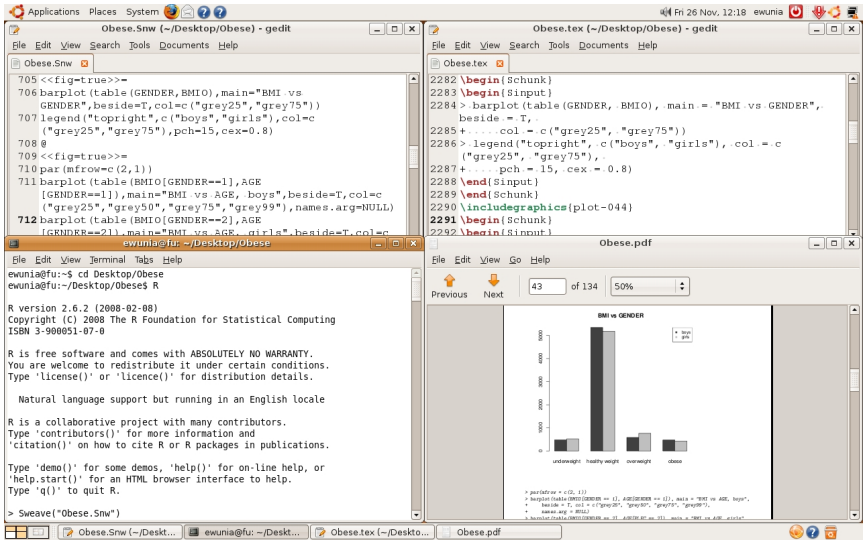

Figure 4. A screenshot of the R package

`www.r-project.org` [14]. Packages of R can be developed using programming languages: S, C/C++, Java.

User interface and formal language The syntax of language R is very similar to that known from the S or S-PLUS languages, which are previous versions of R [15, 16]. This is an interpreted language, hence it is possible to write scripts using functions from C/C++ libraries or from Java packages. The number of packages for R still grows, and there are numerous packages dedicated for various domains of science which applies statistical computations, e.g. *Analysis of Ecological Data*, *Cryptographic Boolean Functions*, or *Chemical Thermodynamics and Activity Diagrams*, cf. [17].

R is accessible for Linux, Windows and MacOS operating systems. The basic set of packages takes about 200 MB of disk space. R is free and under General Public License (GNU GPL), cf. [14, 5]. The R package can be used in Windows systems, as far as in Linux, with different GUIs. In particular, R can be explored in one of the four following ways:

1. Interactive mode – commands are entered using the console mode and executed immediately. Results are displayed in a console, too.

2. Batch mode – scripts (batch files) with commands and/or source code in the R language are created using text editors and executed via main menu of the editor, e.g. a very popular Gnome text editor named "gedit" and its plugins, cf. [18]. Results of computations are not displayed until the `print()` command is used.

3. IDE mode, Integrated Development Environment mode – scripts are edited using environments for R, e.g. Rcmdr [19], RKWard [20], Tinn-R [21], or StatET – a plugin for Eclipse [22]. Results are displayed also by adequate functions of IDEs.

4. Noweb files [23] – a toolkit for the so-called *literate programming* (see the paragraph below) that makes it possible to combine the R source code (the so-called "code chunks") with adequate documentation in LaTeX (the so-called "text chunks") in the same file.

Literate programming is a technique of programming based on putting *code chunks*, here: the code in R, to the text formatted with LaTeX The idea of literate programming has been adopted for R in 1994 by Ramsey[1] [23, 24].

In particular, in computations presented in this paper, the Sweave package is used for exploring the idea of literate programming [25, 26]. Sweave is a toolkit – a collection of AWK scripts and shell scripts used for entering R source codes to LaTeX documents (see Figure 4). Types of files in which Sweave can be applied, are described by the following extensions: `.Rtex`, `.Stex` – LaTeX files or `.Rnw`, `.Snw` – text files.the following operator enable entering *code chunks* while programming with R: `'«Name, Options»='`. This code is possible to be reused, with the same operator. *Documentation chanks* begin with '@'. Sample options related to chanks in an output file, are enumerated now:

1. `echo=false` – no text is saved to code chunk,

2. `results=hide` – no evaluated results are saved to code chunk,

3. `fig=true` – a chart is created and placed in the document as `.pdf` or `.eps`,

4. `split=true` – the result of code chunk is placed in a separated file.

[1] It is worth adding that "code chunks" can be applied to another programming languages e.g. SAS, Java, etc. when programs are documented using LaTeX .

Table 1. Commands and functions for statistical tests: a comparison within presented software

Test of Independence χ^2	Gnumeric	`Tools; Statistical Analysis; Test of Independence`
	Octave	`chisquere_test_independence()`
	R	`Chisq.test()`
	Statistica	`Statistics; Basic Statistics/Tables; Tabels and Banners`
Tests of Proportions	Gnumeric	—
	GNU Octave	`prop_test_2()`
	R	`prop.test()`
	Statistica	`Statistics; Basic Statistics/Tables; Difference tests: r, %, means`
Normality Tests	Gnumeric	`Tools; Statistical Analysis; Normality Tests`
	GNU Octave	`kolomogorov_smirnov_test()`
	R	`Shapiro.test(), cvm.test(), lillie.test()`
	Statistica	`Statistics; Basic Statistics/Tables; Frequency Tables`
t-tests	Gnumeric	`Tools; Statistical Analysis; Two means`
	GNU Octave	`t_test_2(), welch_test(), z_test_1()`
	R	`t.test()`
	Statistica	`Statistics; Basic Statistics and Tables; t-test`
Nonparametric Tests	Gnumeric	`Tools; Statistical Analysis; Two medians sign test`
	GNU Octave	`wilcoxon_test(), u_test()`
	R	`Wilcox.test()`
	Statistica	`Statistics; Nonparametrics; Comparing Two Independent Samples`

A source file containing code in R (code chunks) and LaTeX (text chunks) requires to be compiled twice (at least):

`.Snw` or `Rnw` file with code chunks

↓ 1. *compiling code chunks in R*

`.tex` source & results of compiling the R code

↓ 2. *compiling the output file*

`.dvi` or `.pdf` output file with results of executing code chunks

Table 2. Commands or functions for charts: a comparison of presented software

Bar Plot	Gnumeric	`Insert; Chart; Column`
	GNU Octave	`bar(), plot()`
	R	`barplot(), plot()`
	Statistica	`Graphs; Graphs 2W; Bar/Column Plots`
Histogram	Gnumeric	`Insert; Chart; Histogram`
	GNU Octave	`hist()`
	R	`hist()`
	Statistica	`Graphs; Graphs 2W; Histograms`
Box Plot	Gnumeric	`Insert; Chart; BoxPlot`
	GNU Octave	`errorbar()`
	R	`boxplot()`
	Statistica	`Graphs; Graphs 2W; Boxplots`

where "↓ 1. . . ." and "↓ 2. . . ." denote the process of the first and the second compilation of the source code in LaTeX, respectively. The first compilation resolves code chunks in R into figures, charts, tables, etc.). The second compilation creates an output file for the text and all the so-called *floating bodies* i.e. tables, charts, figures that come from executing the code in R; the file can be `.dvi`, `.pdf`, `.ps` etc. Sample usage of code chunks in R and Sweave are given in Algorithms 1-3.

R and other free packages for applied statistics: a comparison Collection of commands that perform basic statistical tests in Gnumeric, Statistica, GNU Octave and R, is presented in Table 1. Commands for chart presentation of computation results in the described software are collected in Table 2.

The presented software were compared with respect to the following criteria: accessibility, computation speed, and scalability. As it can be seen in Table 3, we especially underline the advantages of R package and language, which seems to be the most universal scalable free and open package.

4. A case study: application of R and Sweave to biometrics

4.1. Dataset

The analysed dataset describes results of questionnaire (polled by Municipality of the City of Bytom, Poland) examination of health status of pupils of primary

Table 3. Comparison of features of presented software

	Gnumeric	**Statistica trial**	**GNU Octave**	**R**
GUI	YES	YES	NO	YES
License	GNU GPL	30 days trial	GNU GPL	GNU GPL
Required OS	Windows Linux	Windows	Windows Linux	Windows Linux
Speed of computations	low	medium	high	high
Additional packages or plugins	NO	YES	YES	YES

and secondary schools: children and adolescents in 12 age groups: 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18 years, in the 2003/2004 school year. The dataset contains 13 778 elements, including 6 879 boys and 6 899 girls. The number 13 778 represents the population very well, because the total number of pupils in Bytom was 20 426 (2003/2004). In particular, anthropometric data on body height (in cm), body weight (in kg), and age (in years) were collected and taken into account in the presented examples and computations.

4.2. Methods

The so-called *Body Mass Index*, (BMI, $[kg/m^2]$) is computed from body weight and body height as a descriptive characteristic of each pupil [27].

$$\text{BMI} = \frac{\text{weight}}{\text{height} * \text{height}} \tag{1}$$

In medical examination and analysis, adults (above the age of 18) are qualified as "underweight" when BMI $\leq$ 20 kg/m^2, as "healthy weight" when 20 $\leq$ BMI $\leq$ 25, as "overweight" when 25 $\leq$ BMI $\leq$ 30, and as "obese" when BMI $\geq$ 30. However, for children and adolescents, the so-called *centile ranks* and *growth charts* worked out by Institute of Mother and Child in Warsaw, Poland [28]. The suspicion of underweight is found if the BMI is placed below the 5th percentile, overweight – if BMI varies between the 85th and 95th percentile, and obesity if BMI exceeds the 95th percentile [29].

The following statistical tests are applied on the data:

- Non-parametric tests – Shapiro-Wilk normality test, Wilcoxon rank-sum test,
- Parametric tests – Student's t-test.

p-values smaller than 0.05 were considered statistically significant.

4.3. Computations and codes

This section contains three examples that illustrate commonly used computational methods of statistical analysis provided by the R language and package Sweave. The computations are run on datasets characterized in Section 4.1. In particular, the emphasis is put on codes in R and its processing to results, charts and/or figures that support statistical analysis. The idea of literate programming (see Section 3) is taken into account, too.

Example 1 *Sample sizes (body weight and body height) of pupils in 12 age groups are analysed. The code in R that generates results of the analysis is given in Algorithm 1. The structure of body weight and body height in the age groups are illustrated by charts given in Figure 5.*

Algorithm 1: The code in R that generates analysis depicted in Figure 5

```
<<fig=true>>=
# Figure 5. (Chart 5.)
par(mfrow=c(2,2),mar=c(3,2,2,1))
boxplot(HEIGHT~AGE,cex.axis=0.6,main="(a) Box plot of HEIGHT")
boxplot(WEIGHT~AGE,cex.axis=0.6,main="(b) Box plot of WEIGHT")
barplot(table(AGE,HEIGHTG),xlab="HEIGHT",col=rainbow(12),
cex.axis=0.6,cex=0.6)
legend("topright",c("7 years","8 years","9 years","10 years","11 years",
"12 years","13 years","14 years","15 years","16 years","17 years",
"18 years"),lwd=5,col=rainbow(12),ncol=2,cex=0.6)
barplot(table(AGE,WEIGHTG),xlab="Weight",col=rainbow(12),
cex.axis=0.6,cex=0.6)
@
```

Example 2 *The measurable continuous variables: body weight and body height are described with descriptive characteristics: mean $\bar{x}$ and standard deviation $S, \bar{x} \pm S$. Student's t-test is applied to compare average values of body weight*

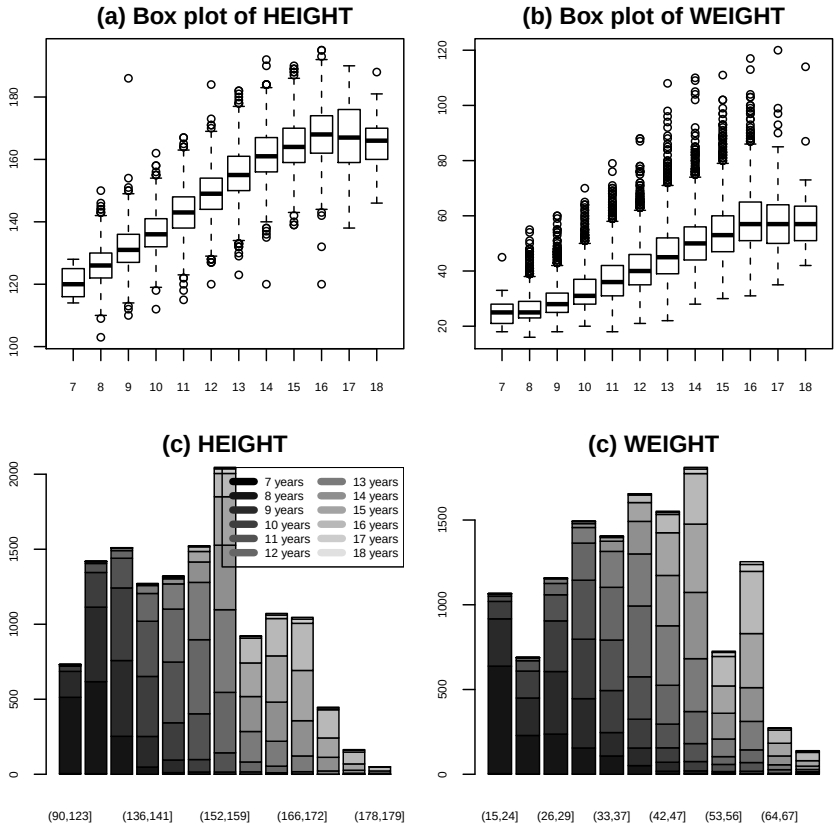

Figure 5. Dataset analysis generated by the code in Algorithm 1: boxplots of mean body height and body weight (on the top), and histograms of body height and body weight in age groups (on the bottom)

and body height, in two independent samples (in two genders), assuming that data come from normal distribution (the Shapiro-Wilk test). When this assumption is not satisfied, the non-parametric test is applied – Wilcoxon rank-sum test, "Mann-Whitney" test. The code that enables this analysis in R, is given as Algorithm 2.

The analysis of average body height and body weight shows that the highest increase of body height of boys is for 9-14 years, when pubertal spurt is noted, and

Algorithm 2: The code in R that generate analysis presented in Table 4

```
<<>>=
# Table 4.
for(i in 7:18)
{cat("AGE:", i, "lat","\n")
V1<-HEIGHT; #V1<-WEIGHT
A<-na.omit(V1[AGE==i & GENDER==1])
B<-na.omit(V1[AGE==i & GENDER==2])
cat("N:","BOYS-",length(A),"GIRLS-",length(B))
p1<-shapiro.test(A)$p.value
p2<-shapiro.test(B)$p.value
ifelse(p1>=0.05 && p2>=0.05,print(t.test(A,B)),
print(wilcox.test(A,B)))
print(paste("BOYS-",mean(A),sd(A)))
print(paste("GIRLS-",mean(B),sd(B)))}
@
```

Table 4. Dataset analysis generated by the code in Algorithm 2: measures of body weight and body height of boys and girls in age groups (Significance $*p < 0.05$, t-test, Wilcoxon rank-sum test – "Mann-Whitney" test).

	Boys			Girls			*p*-value	
Age	*N*	Height $\bar{x} \pm S$	Weight $\bar{x} \pm S$	*N*	Height $\bar{x} \pm S$	Weight $\bar{x} \pm S$	Height	Weight
7	1	128.0	28.0	8	119.9±5.2	26.5±8.8	-	-
8	744	126.4±6.0	26.9±5.8	690	125.3±6.0	26.1±5.6	<0.001*	<0.01*
9	700	132.0±6.6	30.2±6.4	779	131.0±6.6	29.4±6.5	<0.001*	<0.01*
10	757	137.1±6.7	33.5±7.8	738	136.1±6.7	32.4±7.1	<0.001*	<0.05*
11	815	143.3±7.5	38.1±9.4	684	142.8±7.4	36.9±9.0	0.31	<0.05*
12	808	148.8±7.3	41.3±9.5	818	149.4±7.5	41.3±9.4	<0.05*	0.67
13	811	155.1±9.1	46.6±11.5	869	155.4±7.3	46.6±10.0	0.17	0.14
14	749	162.9±9.2	52.5±12.0	735	160.0±6.6	50.2±9.2	<0.0001*	<0.01*
15	732	167.7±8.5	56.6±11.9	787	161.6±6.1	52.8±9.2	<0.0001*	<0.0001*
16	651	173.6±7.5	62.7±11.8	701	163.4±6.4	55.8±9.7	<0.0001*	<0.0001*
17	83	173.2±8.9	62.4±13.4	67	160.2±6.9	53.8±8.7	<0.0001*	<0.0001*
18	28	170.3±8.0	60.5±13.3	23	160.8±6.6	55.5±10.1	<0.0001*	0.06

of girls – 8-13 years. In the case of body weight no significant increments for both genders can be observed. Statistically significant differences of body height due to gender are observed for the following age groups: 8-10 years, 12, and 14-18 years. Statistically significant differences of body weight are observed for the following:

8-11 years, and 14-17 years. see Table 4.

Example 3 *Frequencies and percentiles are used to analyse unmeasurable data (*underweight, healthy weight, overweight, obesity*) presented in contingency table. The histograms are used to show a frequency distribution. The so-called* pie-chart *illustrate the percentiles of both samples, see Figure 6. The code that enables this analysis in R, is given as Algorithm 3.*

The conclusions from computations shown in Example 3 are as follows: due to the established intervals for BMI values [28], 7.3% of children and adolescent may suffer from underweight, boys – 6.7% and, similarly, girls – 7.9%, about 77% children and adolescent are of the healthy weight. About 10% are overweighted and 6.4% are obese, see Figure 6(d). The fraction of boys with overweight or obesity is 14.1% (971 of 6 879) and of girls – 17.3% (1 196 of 6 899) (Figure 6a). These results are consistent with those from 2005, which shows that from 16 to 22% of children 7 to 17 years old inhabiting European countries, are overweighted or obese, and among them 4-6% are obese [27].

5. Conclusions

The paper presents popular free and/or open software packages dedicated for statistical computations and applied statistics. In particular, four packages, i.e. Gnumeric, Statistica Trial, GNU Octave and R are described. Popular statistical tests are examined in each package and commented in Table 1. Commonly used charts are tested and adequate commands are collected in Table 2. A comparison of these packages is done, mostly from the point of view of their scalability and portability, speed of computations, and additional packages or accessible plugins, see Table 3.

Especially, the authors focus on one of the most popular free and open software package for applied statistics: R. It is selected as possibly the most versatile and scalable software for applied statistics. Features of R and associated packages/plugins/environments are explored, described and critically analysed in Section 3. A case study in biometrics is made in Section 4; an analysis of sample sizes and descriptive characteristics of pupils in 12 age groups according to gender is made using the functionality of the described R package.

The presented analysis and comparison can be found as a handy and useful

Algorithm 3: The code in R that generates analysis depicted in Figure 6

```
<<fig=true>>=
#Figure 6. (Chart 6.)
BMI=WEIGHT/((HEIGHT/100)^2)
#Sign to BMI category based on age/sex charts.
BMI7<-BMI[AGE==7]
summary(BMI7)
BMIO7C=cut(BMI7[AGE==1],c(0,13.6,17.3,19.0,50),
c("underweight","healthy weight", "overweight", "obese"))
BMIO7D=cut(BMI7[GENDER==2],c(0,13.4,18.0,20.2,50),
c("underweight","healthy weight", "overweight", "obese"))
BMI7<-c(BMIO7C,BMIO7D)
length(BMI7) (...)
BMIO=c(BMI7,BMI8,BMI9,BMI10,BMI11,BMI12,BMI13, BMI14,BMI15,
BMI16,BMI17,BMI18)
BMIO<-factor(BMIO)
attr(BMIO,"levels")<-
c("underweight","healthy weight", "overweight", "obese")
#Figure 6. (Chart 6.).
par(mfrow=c(2,2),mar=c(3,2,2,1))
T<-table(GENDER,BMIO)
balloonplot(T,text.size=0.6,main="(a) BMI vs GENDER",
dotcolor="grey",label.size=0.6,show.zeros=TRUE,
show.margins=TRUE,rowmar=2,colmar=0.6)
barplot(table(BMIO[GENDER==1],AGE[GENDER==1]),
main="(b) BMI vs AGE, boys 1",beside=T,col=
c("grey25","grey50","grey75","grey99"),
names.arg=NULL)
barplot(table(BMIO[GENDER==2],AGE[GENDER==2]),
main="(c) BMI vs AGE, girls 2",beside=T,col=
c("grey25","grey50","grey75","grey99"))
legend("topright",c("underweight","healthy weight",
"overweight","obese"),col=c("grey25","grey50",
"grey75","grey99"),pch=15,cex=0.6)
x=summary(BMIO)
percent=format(100*x/sum(x), digits=2)
pie(x, init.angle=100, cex=0.6,main="(d) BMI",
labels=paste(names(x), percent, "%"),
col=c("grey20","grey40","grey60","grey80","grey99"))@
```

introduction to R in particular and to free and open software packages for applied statistics in general.

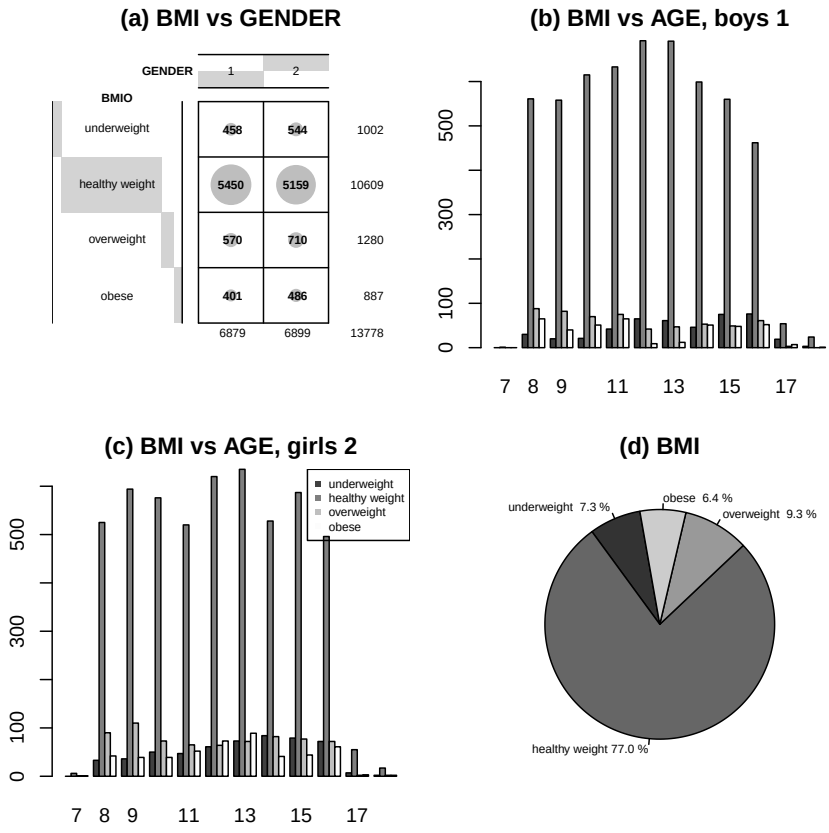

Figure 6. Dataset analysis generated by the code in Algorithm 3: (a) Summary of BMI according to gender, (b) histogram of BMI according to age (boys), (c) histogram of BMI according to age (girls), (d) Percentage of underweight, healthy weight, overweight, and obesity by BMI

References

[1] `http://www.mathworks.com/products/matlab` (access: June 20, 2011).

[2] `http://www.spss.pl` (access: June 20, 2011).

[3] http://www.sas.com (access: June 20, 2011).

[4] http://www.minitab.com (access: June 20, 2011).

[5] http://www.r-project.org (access: June 20, 2011).

[6] http://www.gnome.org/projects/gnumeric (access: June 20, 2011).

[7] http://git.gnome.org/browse/gnumeric (access: June 20, 2011).

[8] http://www.statsoft.pl (access: June 20, 2011).

[9] Eaton, J. W. and Bateman, D., *GNU Octave. Edition 3 for Octave version 3.2.3*, Free Software Foundation, Boston, 2007.

[10] http://www.gnu.org/software/octave (access: June 20, 2011).

[11] http://www.openepi.com (access: June 20, 2011).

[12] http://www.gnu.org/software/pspp/ (access: June 20, 2011).

[13] http://statpages.org/javasta2.html (access: June 20, 2011).

[14] http://cran.r-project.org (access: June 20, 2011).

[15] Chambers, J., *Programming with Data: A Guide to the S Language*, Springer, 1998.

[16] Rabe-Hesketh, S. and Everitt, B., *Analyzing Medical Data Using S-PLUS*, Springer-Verlag New York, 2001.

[17] http://r.meteo.uni.wroc.pl (access: June 20, 2011).

[18] http://projects.gnome.org/gedit/ (access: June 20, 2011).

[19] http://cran.r-project.org/web/packages/Rcmdr (access: June 20, 2011).

[20] http://rkward.sourceforge.net (access: June 20, 2011).

[21] http://cran.r-project.org/web/packages/TinnR (access: June 20, 2011).

[22] http://www.walware.de/goto/statet (access: June 20, 2011).

[23] Ramsey, N., *Literate programming simplified*, IEEE Software, Vol. 11, No. 5, 1994, pp. 97–105.

[24] Knuth, D. E., *Literate Programming*, California: Stanford University, 1992.

[25] Leisch, F., *Sweave User Manual*, `www.ci.tuwien.ac.at/leisch/Sweave`, `(access: June 20, 2011)`.

[26] Lenth, R. V. and Hojsgaard, S., *SASweave: Literate Programing Using SAS*, Journal of Statistical Software, Vol. 19, No. 8, 2007, `http://www.jstatsoft.org, (access: June 20, 2011)`.

[27] Lobstein, T., Rigloy, N., and Leach, R., *International obesity task force*, The International Association for the Study of Obesity, Brussels, 2005.

[28] Matusik, P., Małecka-Tendera, E., and Nowak, A., *Methods used in paediatric practice for nutritional status estimation in children (Metody stosowane w praktyce pediatrycznej do oceny stopnia odżywienia dzieci)*, Endokrynologia, otyłość i zaburzenia przemiany materii, Vol. 1, No. 2, 2005, pp. 6–11, (in Polish).

[29] *Computation of centiles and z-scores for height-for-age, weight-for-age and bmi-for-age*, `www.who.int/growthref/en`, 2007 (access: June 20, 2011).