ISSN 2286-4822
www.euacademic.org

EUROPEAN ACADEMIC RESEARCH
Vol. II, Issue 7/ October 2014

Impact Factor: 3.1 (UIF)
DRJI Value: 5.9 (B+)

An Introduction to Statistical Software Usages in Food and Nutrition Research: A Review

MD RAMIM TANVER RAHMAN[1]
ZAIXIANG LOU
WENPU CHEN
XIANG LI
RIYADH A. S. THABIT
State Key Laboratory of Food Science and Technology
School of Food Science and Technology
Jiangnan University, Wuxi
P.R. China
ABDULLAH-AL-MARUF
Department of Livestock Services
Ministry of Fisheries and Livestock
The Peoples' Republic of Bangladesh
TUNAJJINA RAHMAN
Department of Economics
Jagannath University, Dhaka
The Peoples' Republic of Bangladesh
AZHARI SIDDEEG
Department of Food Science and Technology
Faculty of Engineering and Technology
University of Gezira, Wad Medani
Sudan

Abstract:

In the past decade, there has been an explosion of high-content and high-throughput data associated with a large number of disease states, chemical exposures, biological species, food and nutritional researches. To fully interpret this information, it has become necessary to develop a range of software tools that will identify the potentially biological patterns within given sets of data, and present- it in a

---

[1] Corresponding author: ramimbau@yahoo.com

context that is both understandable to non specialists, and searchable so that the data underlying the constructed networks can be reviewed and assessed. To this end, a number of analysis tools have been developed, and this paper will provide a short overview of the most commonly used approaches of both commercial and open access software suites for food and nutrition researches.

Key words: Statistical Software Usages, Food and Nutrition Research

Introduction

Food and nutrition science is a multidisciplinary subject, typical curricula including chemistry, biochemistry, microbiology, physics, marketing, business studies as well as some mathematical input (Rahman et al., 2014). The practicing food and nutrition scientist, whether in the laboratory, on the production floor or in the marketing office has the task of interpreting data and making important decisions on his/her conclusions. It is here that the subject of statistics enters. While many undergraduate scientists receive at least one statistics course the instillation of statistical appreciation may be long in the attainment. Indeed, the UK STEPS project (statistical education through problem solving) recognizes problems of “motivation” and “a wide spread of numerical skills” when teaching statistics as a support subject to non-mathematicians. Even later, when established in industry, some food scientists may be weary and reluctant about using statistics or attempting the statistical approach to experimentation (John, 1995). That is why today the software is a ubiquitous and essential tool in the modern measurement and testing laboratory. This of course is said tongue in cheek but it does highlight an important class of problems associated with software viz. the difficulties that can arise when the user is able to modify the way in which a program operates. User customization of software is an area of increasing concern in laboratories seeking to acquire and/or maintain accreditation to internationally recognized quality standards such as ISO 17025 (formerly ISO Guide 25) or GLP/GMP (Hardcastle, 2000). Statistical procedures can be applied to several areas within food science: sensory evaluation; instrumental measures of quality; correlation of sensory and instrumental measures; chemical analysis; statistical quality control; comparison of nutritional effects of processing. All of the above involve gathering data or making measurements of some kind. Science is based on recording measurements and within food science there can be a wide range of different types. As can be seen, statistics deals with the "collecting, organizing and analyzing" of data, usually in the form of numbers as above. The particular feature being measured or observed is described as a variable. Each measurement or observation is subject to variation owing to one or more factors, random or otherwise, some of which contribute to error. Measurements can be single or replicate depending on the circumstances. Statistical methods can also be employed in the way in which the data are obtained ("statistical sampling"). Let us envisage a typical "data confrontation" situation in which the scientist finds her/himself – the data may be in a report or publication or may have been generated by the scientist. The problem is one of deciding several things: What do these data mean? Are the data valid? Are they convincing enough to progress on to more work? To provide answers to these questions, and to aid statistical appreciation, a step-wise approach is recommended (John, 1995). Along with the wide application of various automated analytical instrument in the medical professional field and the widespread penetration of electronic computer technology to the medical department, the food nutritional statistical and analysis has entered into a new era with the main feature of automation, information and network. Although the statistical and analysis automation of medical nutrition is only a dream at present, the nutritional statistical and analysis system (referred to as NSAS, the same as below) composed by the network and information computer system has entered into a booming period and the arrival of computer statistical analysis instead of the computer has become an indisputable fact (Hardcastle, 2000). The computer automation statistical analysis system can deal with the rapid-growth statistical data, it can fully play the advantages of high-speed computing calculation and accurate statistical data, ease the contradiction between the high speed determined by the automatic test instrument and computer system and manual report result and the low efficiency, thus making a great contribution to provide a large amount of data for the nutrition statistical analysis. Recently, some companies have begun to develop some systems as NSAS.

Currently, the statistical analysis of scientific data is basically dependent on the computer, the reform of such kind of data processing means quietly rose at the late eighties in the country and there are various statistical analysis software's and procedures compiled by each research unit. As times go on, the majority of statistical software's are gradually eliminated due to the backward algorithm, poor data format specification, disorder output format and lack of integrity (Margolis et al., 2011). In recent days, large-scale statistical analysis systems or software's are accepted by the majority of research workers, especially the researchers in the statistical community, such as SAS, SPSS, BMDP, GLIM, STATA and EGRET. However, this software's are some common statistical analysis software's, which cannot be well adapted by the professional statistical analysis format in food industry (Church and Krines, 2008).Therefore, the set of nutrition statistical analysis software was developed. Although there are many aspects in the system to be improved, it is hoped that the system can explore a new path for the development of medical field informatization (Koroušic et al., 2013 ; Pereira et al., 2009).

Function of Food and Nutrition Statistical and Analysis System:

Data acquisition collects the individual's (needing to investigate the nutrition status) food intake amount within a certain time and type-in the collected data into the database. Vladislav et al. (2011) shows the section requires the people to participate and the system needs to provide the entry personnel operation interface. Data query The material needed to be queried in the section can be divided into three categories: the first one is providing the enter data for the professional to review and audit the reasonableness of each group of data; the second one is querying the nutrition composition content in each food; the third one is querying the result after the software statistical analysis for the expert to offer the reasonable evaluation on the nutrition health condition of test subject (Ouyang, 2013). Nutritional data calculation is the fundamental basis for the nutrition expert to evaluate the nutrition status, the function requires calculating the entry original data according to the professional formula on the nutrition, obtaining the classified intake amount of each nutrition element and providing the various display modes and printing function. Data analysis is an important function in the software. After obtaining the calculation result, the calculation result shall be saved in the database for spare. With the contrast with the standard data, it can obtain the result, analyze different nutrition status from different result and it shall visually display in the forms of chart and text.

System module design:

The system adopts the modular design program and it can be mainly divided into two modules of maintenance system and using system. Each independent part is controlled by the master module, wherein, the module in the using system mainly includes the nutrition calculator, dietary nutrition evaluation and healthy weight scale (Fig. 1).

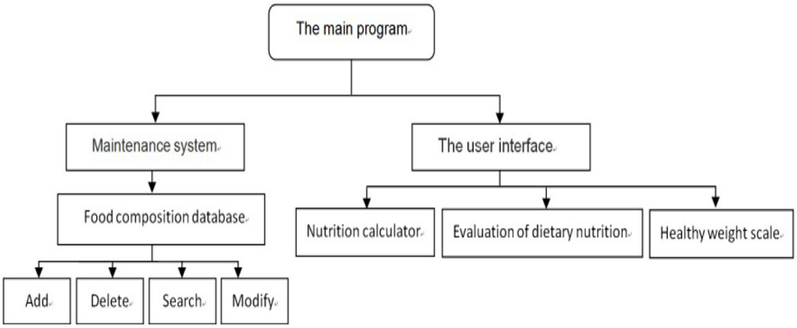


Fig 1. System data flow diagram (general)

Nutrition calculator is the basic calculation statistical tool in the system. It can calculate the contained nutrition component in these foods through the food intake amount within a certain time and store the calculation result in the database and conduct the statistical analysis on the stored data. The calculated data includes the energy, protein, fat, carbohydrate, dietary fiber and various vitamins, trace elements and others in total of 28 (Peilong et al., 2013). For example, if we want to analyze the nutrition status of the sampling population in a region, firstly input the intake food variety and amount within a certain time and the system can automatically calculate the nutrients content according to the input and make the basic evaluation on the overall nutrition condition and conduct the statistical operations of sum, average, variance, chi-square test through NSAS statistical function according to the demand. The system gives full play to the advantages of high calculating speed and high calculation accuracy, thus simplifying the work of nutrition researcher.

Dietary nutrition assessment module is used to analyze and evaluate different respondents' nutrition status. In the module, the database is classified according to the investigation group class and meal species. When applying, it can find the corresponding nutrition component and content according to the specific group, time, intake food class and amount and conduct the statistical analysis. The operator shall firstly select which special group the respondent belongs to, such as infant, children, pregnant women, nursing mother, adult and people with illness, then input the respondent's age, gender, height, weight and physical type and other basic conditions and the intake food class and amount within a certain time, the analysis system will calculate the corresponding nutrition component class and intake amount according to the input condition and make contrast with the national formulated standard intake amount and further make the reasonable evaluation on the respondent's nutrition condition. In addition to be used for the nutrition assessment on the patient in the hospital, the module can be used for the healthy person to make the nutrition assessment, such as infant and children, thus providing the objective basis for the dietitian and professional as the reference for the nutrition counseling guidance and treatment to help the respondent establish healthy diet structure and good eating habit.

Healthy weight scale is a module to propose different nutrition improvement program according to different people's physical conditions by use of computer's artificial intelligence technique.

According to the inputted height, weight and other physiological conditions, it can make the analysis, judge the nutrition status and offer the improvement program. The module also has the nutrition catering function and the dietician can scientifically cater for the patient with the function. According to patient's nutrition status, the system will firstly analyze the needed nutrition component, offer the corresponding food variety table, then the dietician selects the food in the table, the system will conduct the real-time calculation on the difference between the intake amount of nutrition component and needed nutrition component, then the dietician can formulate a reasonable diet for patient through increasing or decreasing the food variety or intake amount according to the difference. It can not only be used for the hospital to cater for the patient, but also in the school, restaurant, hotel and other catering industries.

System Test and Experiment

The system function test adopts the remote access service procedure; user can open the browser at any computer accessible to the Internet and log in to the system. It tests a day's recipe for a 30-year-old female. The tester's physical condition is as follows: height, 165 cm; weight: 53 kg; engage in the light physical labor; Data entry date: 2013-8-17. A day's recipe for the tester can be seen in Table 1. After the analysis of the software, the obtained results can be shown in Fig. 2. The excessive nutrients are displayed in red, the normal in green and those below the standard in black. After the hand computation for 3 times, the obtained value in the software is in consistent with the calculation result (Du and Yun, 2013).

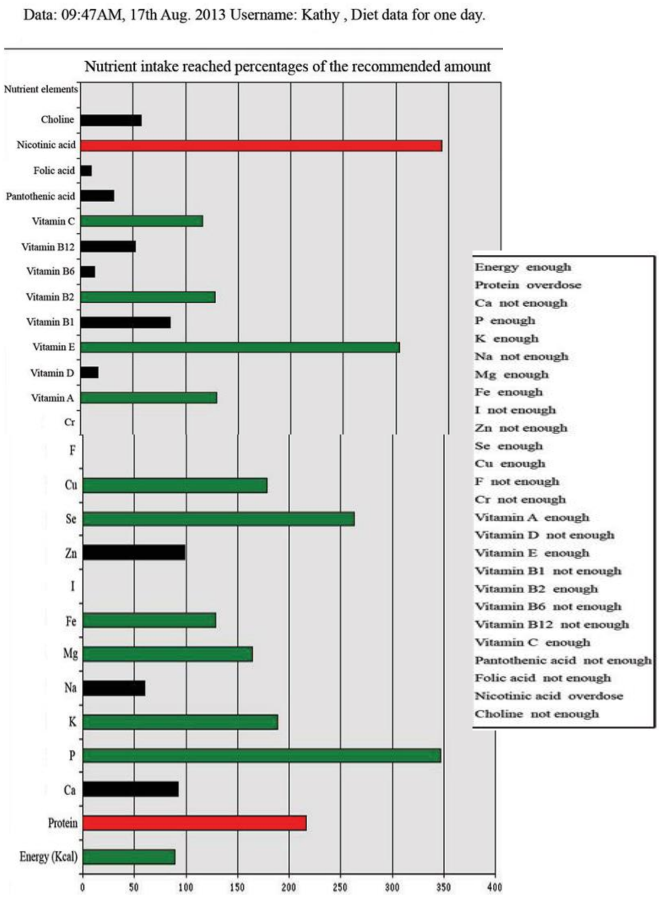
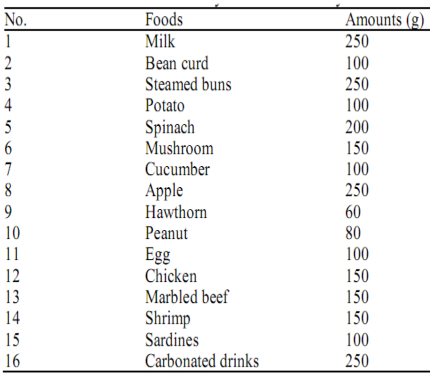


Fig 2. Tester's nutrition status assessment result

Table 1 : Foods amounts taken in by a tester for one day



No.	Foods	Amounts (g)
1	Milk	250
2	Bean curd	100
3	Steamed buns	250
4	Potato	100
5	Spinach	200
6	Mushroom	150
7	Cucumber	100
8	Apple	250
9	Hawthorn	60
10	Peanut	80
11	Egg	100
12	Chicken	150
13	Marbled beef	150
14	Shrimp	150
15	Sardines	100
16	Carbonated drinks	250

SAS:

SAS (Statistical Analysis System) is a software suite developed by SAS Institute for advanced analytics, business intelligence, data management, and predictive analytics. It is the largest market-share holder for advanced analytics. SAS was developed at North Carolina State University from 1966 until 1976, when SAS Institute was incorporated. SAS was further developed in the 1980s and 1990s with the addition of new statistical procedures, additional components and the introduction of JMP. A point-and-click interface was added in version 9 in 2004. A social media analytics product was added in 2010. SAS is a software suite that can mine, alter, manage and retrieve data from a variety of sources and perform statistical analysis on it. SAS provides a graphical point-and-click user interface for non-technical users and more advanced options through the SAS programming language. SAS programs have a DATA step, which retrieves and manipulates data, usually creating a SAS data set, and a PROC step, which analyzes the data. Each step consists of a series of statements. The DATA step has executable statements that result in the software taking an action, and declarative statements that provide instructions to read a data set or alter the data's appearance. The DATA step has two phases, compilation and execution. In the compilation phase, declarative statements are processed and syntax errors are identified. Afterwards, the execution phase processes each executable statement sequentially. Data sets are organized into tables with rows called "observations" and columns called "variables". Additionally, each piece of data has a descriptor and a value.

The PROC step consists of PROC statements that call upon named procedures. Procedures perform analysis and reporting on data sets to produce statistics, analyses and graphics. There are more than 300 procedures and each one contains a substantial body of programming and statistical work.[2] PROC statements can also display results, sort data or perform other operations. SAS Macros are pieces of code or variables that are coded once and referenced to perform repetitive tasks.

SAS data can be published in HTML, PDF, Excel and other formats using the Output Delivery System, which was first introduced in 2007.[7] The SAS Enterprise Guide is SAS' point-and-click interface. It generates code to manipulate data or perform analysis automatically and does not require SAS programming experience to use (Encyclopedia, 2010; SAS, 2010; Lora et al., 2012; Arthur Li. 2013; Buck, 2013).

The SAS software suite has more than 200 components. Some of the SAS components include:

- Base SAS - Basic procedures and data management
- SAS/STAT - Statistical analysis
- SAS/GRAPH - Graphics and presentation
- SAS/OR - Operations research
- SAS/ETS - Econometrics and Time Series Analysis
- SAS/IML - Interactive matrix language
- SAS/AF - Applications facility
- SAS/QC - Quality control
- SAS/INSIGHT - Data mining
- SAS/PH - Clinical trial analysis
- Enterprise Miner - data mining

Online Resources:

Online help: Type help in the SAS display manager input windows; Sample Programs, distributed with SAS on all platforms.

SAS Institute Home Page: http://www.sas.com; SAS Institute Technical Support: http://support.sas.com/resources/; Searchable index to SAS-L, the SAS mailing list: http://www.listserv.uga.edu/archives/sas-l.html; Usenet Newsgroup (equivalent to SAS-L): comp.soft-sys.sas; Michael Friendly's Guide to SAS Resources on the Internet http://www.math.yorku.ca/SCS/StatResource.html#SAS; Brian Yandell's Introduction to SAS: http://www.stat.wisc.edu/~yandell/software/sas/intro.html

Resources: Introductory Books; Mastering the SAS System, 2nd Edition, by Jay A. Jaffe, Van Nostrand Reinhold, Quick Start to Data Analysis with SAS, by Frank C. DiIorio and Kenneth A. Hardy, Duxbury Press.; How SAS works: a comprehensive introduction to the SAS System, by P.A. Herzberg, Springer Verlag; Applied statistics and the SAS programming language, by R.P Cody, North-Holland, New York. The bulk of SAS documentation is available online, at http://support.sas.com/documentation/onlinedoc/index.html. A catalog of printed documentation available from SAS can be found at http://support.sas.com/publishing/index.html (Phil, 1995)

SPSS:

"SPSS is a comprehensive system for analyzing data. SPSS can take data from almost any type of file and use them to generate tabulated reports, charts, and plots of distributions and trends, descriptive statistics, and complex statistical analysis." See www.spss.com for more information. SPSS is the acronym of Statistical Package for the Social Science. SPSS is one of the most popular statistical packages which can perform highly complex data manipulation and analysis with simple instructions. It is designed for both interactive and non-interactive (batch) uses. SPSS Program Functionality: Breadth of functionality: Good SPSS has scores of statistical and mathematical functions, scores statistical procedures, and a very flexible data handling capability. It can read data in almost any format (e.g., numeric, alphanumeric, binary, dollar, date, time formats), and version 6 onwards can read files created using spread sheet/data base software. It also has excellent data manipulation utilities. The following is a brief overview of some of the functionalities of SPSS:

Data transformations Data Examination Descriptive Statistics Cont ingency tables Reliability tests Correlation T - tests ANOVA (Basak et. at., 2014, Jothi et. al. 2014) MANOVA General Linear Model (Release 7.0 and higher) Regression Nonlinear Regression Logistic Regression Loglinear Regression Discriminant Analysis Factor Analysis Cluster anlaysis Multidimensional scaling Probit analysis Forecasting/Time Series Survival analysis Nonparametric analysis Graphics and graphical interface (Argyrous and Levesque, 2007; SPSS, 2006).

BMDP:

BMDP is a statistical package developed in 1965 at UCLA. Based on the older BIMED program, developed in 1960 for biomedical applications, it used keyword parameters in the input instead of fixed-format cards, so the letter P was added to the letters BMD, although the name was later defined as being an abbreviation for Biomedical Package. BMDP was originally distributed for free. It is now offered by Statistical Solutions (Jan, 2009).

GLIM:

GLIM (an acronym for Generalized Linear Interactive Modelling) is a statistical software program for fitting generalized linear models (GLMs). It was developed by the Royal Statistical Society's Working Party on Statistical Computing (later renamed the GLIM Working Party), chaired initially by Nelder(1975). It was first released in 1974 with the last major release, GLIM4, in 1993. GLIM was distributed by the Numerical Algorithms Group (NAG). GLIM was notable for being the first package capable of fitting a wide range of generalized linear models in a unified framework, and for encouraging an interactive, iterative approach to statistical modelling. GLIM used a command-line interface and allowed users to define their own macros. Many articles in academic journals were written about the use of GLIM. GLIM was reviewed in *The American Statistician* in 1994, along with other software for fitting generalized linear models. The GLIMPSE system was later developed to provide a knowledge based front-end for GLIM.[21-27] (RSS, 2007; Francis et al., 1993, GLIM, 2010; Aitkin et al., 1989; Wacholder, 1988; Wolstenholme, 1988)

STATA:

STATA, like SPSS (Statistical Package for Social Science), is a general purpose statistical software package. It is command-based software, and is available for Windows, Macintosh, and Linux systems. Stata provides a highly flexible interactive mode that makes it easier for beginners to learn and use. Stata also supports features for programming and matrix manipulation.

Stata provides a broad range of analyses, including: Descriptive Statistics, Regression models, ANOVA (analysis of variance), Categorical and limited dependent models (e.g., logit and probit), Panel data models, Nonparametric methods, Multivariate methods, Cluster analysis, Survival analysis, Time series analysis.

Advantages:

1) Intuitive data management capabilities. The creation of variables and sub-setting data is simple and straightforward.

2) Wide variety of statistical procedure that can be accessed via a point-and-click method.
3) Syntax is provided so users can learn to code quickly.
4) Users can share complex coding syntax with others.

Possible weaknesses:

1) Not always easy to handle large datasets.
2) Documentation can be sparse for those who want more detail.
3) Complex programming can become challenging. (Youssef Zaghlool, 2012)

EGRET:

EGRET was originally developed at the School of Public Health of University of Washington USA (Mauritsen, 1984). Designed for analysing data from Biomedical and Epidemiology studies, EGRET stands for Epidemiological GRaphics Estimation Testing. It fits generalised linear models with and without random effects and survival models. It concentrates on models for categorical data collected from Epidemiology and Biomedical studies including cohort data, cross-sectional data, case-control data, clinical trial data and survival data. It is widely used by Epidemiologists and Biostatisticians. EGRET for Windows was developed based from an early MS-DOS platform. Released in 1999, the current Window version was developed by a team in CYTEL Software Corporation of Cambridge, MA in the USA.

A comprehensive set of models: Many not available elsewhere:

- Contingency Tables
- Logistic Regression
- Conditional Logistic Regression
- Logistic Regression with Random Effects
- Beta-Binomial Regression
- Poisson Regression
- Weibull Regression Exponential Regression
- Cox Proportional Hazards Regression

- Cox Regression with Time-Dependent Covariates
- Kaplan-Meier Analysis and Plots
- Extensive Post-Fit Analysis with Plots, Including Delta-Betas and Hazard Functions

SiZ 2.0

It is fully validated trial design and simulation package that also analyzes study data. Affordable and easy to learn, SiZ streamlines fixed sample size study planning with more built-in capabilities than any other software.

SiZ is Cytels fully referenced, easily learned design and analysis software for fixed sample size clinical studies featuring:

- A wide array of validated methods in the standard Architect interface, fast for even for first-time users
- A fast trial simulation engine to rapidly and accurately compare designs
- A range of tables, graphs and plots to easily share designs amongst planning team members
- The only trial software that also analyzes data - no more switching applications for data analysis

ToxTools

ToxTools is a complete solution for modeling and testing data from dose-response studies, and for estimating Benchmark Dose Levels (BMDs). The package provides modern, sophisticated modeling and estimation algorithms for fitting dose response data as part of a state-of-the-art suite of graphical and numeric analysis tools for toxicological risk assessment.

With ToxTools, you can choose from a variety of standard dose response model forms for binary and continuous responses. Unique to ToxTools are models for multivariate arrays of mixed discrete/continuous end-points, as well as the ability to accommodate correlated data such as that arising in studies of litters or clusters. Quantitative risk estimation based on the popular Benchmark Dose methodology is easily achieved from any of the available dose response models, including recent methodological advances for defining risk from multiple outcomes. In addition to classical procedures for conducting group comparisons and tests of trend, the software incorporates new methodology for exact trend tests that apply in cases where sample sizes are small. ToxTools incorporates these features in an intuitive, friendly Windows environment that was specifically designed for flexibility and "what if" calculations. It is very easy to conduct multiple analyses at both the modeling and risk estimation stages in order to compare/contrast results. And like all of Cytels products, the package attains the highest quality of commercial standards for reliability, documentation and customer support.

Pathway Analysis Software:

Pathway Identification

DAVID (Database for Annotation, Visualization and Integration Discovery4) is a web based software suite designed to categorize complex, high content, genomic and proteomic data sets (Huang et al., 2009; Shannon et al., 2003). It comprises a set of functional annotation tools allowing scientists to examine biological meaning behind large list of genes, with common outputs from DAVID including identification of over-represented biological pathways; visualisation of gene lists overlaid onto BioCarta and KEGG pathway maps; identification of interacting proteins; identify gene-disease associations.

An advantage of DAVID is its open-source nature and relative rapidity of use. However, this advantage is offset to some degree by the less intuitive output presented by the programme when compared to pathway visualisation tools such as Cytoscape and Ingenuity Pathway Analysis. This concern has been mitigated slightly by the recent publication of a full protocol for use of the software suite (Shannon et al., 2003), although it should still be viewed as a potential limitation for ease of use by the non-specialist.

Pathway Visualisation

Cytoscape

Cytoscape5 is the leading open source software project for integrating biomolecular interaction networks with high-throughput expression data (Cline et al., 2007; Rocke et al., 2009). It uses a graphical front end to allow the visualisation of networks in a cellular context, incorporating expression data from microarray experiments and allowing linkage to online databases such as PubMed. The strongest attribute of Cytoscape is its ability to produce publication quality networks from experimental data, allowing users to visualize not only how different transcripts/proteins links together to form a biological network, but also how the expression levels of these components alter under different experimental conditions through the overlay of single-source array datasets. Such an approach allows both a visual and statistical identification of pathways/networks likely to be impacted by, for example, chemical exposure (Fukushima et al., 2006). Finally, interactive text-mining of databases such as PubMed can be used to annotate and expand networks using published literature.

Ingenuity Pathway Analysis

Ingenuity Pathway Analysis6 is the leading commercial software suite for integrating biomolecular interaction networks with high-throughput expression data. It has been used for a large number of analysis', including liver (Lambert et al., 2009; Gerecke et al., 2009), dermal (Abdel-Aziz et al., 2008) , lung (Naito et al., 2007) and gastric mucosal (Sharma et al. 2009) toxicity. In addition, biological models of disease states

(Gunawardana et al., 2009) and biomarker identification (Committee on toxicity, 2009) have also been investigated using Ingenuity Pathway Analysis Software. As with Cytoscape, the base output of Ingenuity is a generated pathway showing potential network interactions, which is overlayed with links to all datasets/databases used to generate the network. Importantly, multiple datasets can be analysed at one time, for example transcript and proteomic data, increasing the robustness of any conclusions drawn. An additional potential benefit of Ingenuity Pathway Analysis over Cytoscape is the ability to search for potential biomarkers within a generated network. However, it should be noted that this, relatively new, feature has not been sufficiently tested to determine how successful it is at selecting robust biomarkers.

MATLAB and Simulink

MATLAB comes from “matrix laboratory” is a multi-paradigm numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran.

Researchers at biotech and pharmaceutical companies use MathWorks products to perform data analysis and modeling throughout drug discovery, development, clinical trials, and manufacturing. MathWorks software provides a flexible, integrated environment spanning application areas that include: Pharmacokinetic and pharmacodynamic (PK/PD) modeling, Biological image processing, Biostatistics, Bioinformatics, Biochemical modeling and systems biology, Process analytical technology (PAT), Pharmaceutical researchers at any level of programming expertise can perform analysis, visualization, and modeling through both graphical and command-line functionality. Scripting facilities enable straightforward automation of computational workflows.

A) Analyze, visualize, and model biological data and systems

Computational biologists use MathWorks products to understand and predict biological behavior using data analysis and mathematical modeling. MathWorks products provide a single, integrated environment to support pharmacokinetics (PK), bioinformatics, systems biology, bioimage processing, and biostatistics.

MathWorks computational biology products can also be used to: Import, analyze, and model data, and share results, Automate workflow elements, Customize algorithms and tools critical to developing innovative methods for working with unexplored research areas, Leverage proven, commercially supported algorithms and tools etc.

B) Data Analysis for Computational Biology

By providing an integrated environment for computational biology, MathWorks products eliminate the need to work with separate, incompatible tools for import, analysis, and results sharing. It performs a spectrum of analyses including nonlinear mixed-effects, sequence, microarray, phylogenetic tree, mass spectrometry, and gene ontology, Import data from multiple sources, such as databases, file formats, or hardware, Share results with automatically generated HTML reports, data visualizations, or stand-alone tools, Parallelize data analysis to decrease computation time, Automate analyses to implement batch processing of contiguous processes, Mathematical Modeling for Computational Biology.

MathWorks products provide a unified environment for various types of modeling, such as pharmacokinetics (PK) and systems biology. The integrated environment allows you to create and analyze a model to predict and study characteristics of your biological system. Graphical and programmatic modeling tools in SimBiology enable us to analyze and visualize external data, identify a PK model from the PK Wizard, and fit your model to the external data, Implement a reaction network to study the dynamics of a system, Combine models such as PK with a biochemical network, Share results with an automatically generated report etc.

Limitations of Analysis Software:

Although all of the abovementioned approaches will provide a wealth of information, it is important to realize that a number of limitations exist to the analysis, all of which need to be taken into account when interpreting any given analysis.

Database Description link

Drug bang	Drug Target Interactions	http://www.drugbank.ca/
LIGAND	Small Molecular Target Interactions	http://ligand.info/
Entrez OMIM	Human Genes And Genetic Phenotypes	http://ncbi.nlm.nih.gov/omim
KEGG	Kyoto Encyclopedia of Genes and Genomes	http://www.genome.jp/kegg/
db SNP	Single Nucleotide Polymorphism Database	http://www.ncbi.nlm.nih.gov/ projects/SNP/
miRBase	miRNA sequence and targets	http://www.mirbase.org/
Argonaute2	miRNA sequence and targets	http://www.ma.uni-helberg.de/apps/zmf/argonaute/
UniPort	Protein Sequence And Function	http://www.uniprot.org/
HMDB	Human Metabolome Database	http://www.hmdb.ca/
IPI	International Protein Index	http://www.ebi.ac.uk/IPI/IPI help.html
ClinicITrials.gov	Clinical trial information over 170 countries	http://www.clinicaltrials.gov/

Bio Carta	Repository of Graphic respiration of pathways	http://www.biocarta.com/Default.aspx
BIND	Biomolecular Interaction Network Database	http://www.bind.ca
MINT	Molecular interaction Database	http://mint.bio.uninomaz.it/mint.welcome.do
Entrez PubMed	Published Scientific Literature	http://ncbi/nlm/nih.gov/pubmed/

Software Source and Links:
http://www.statease.com/dx9.html#description;
http://catalogs.mhhe.com/mhhe/home.do;
http://www.nmschembio.org.uk/GenericHub.aspx?m=475;
http://www.lgcgroup.com/ etc.

Conclusion:

It is generally acknowledged that the most important changes in statistics in the last 50 years are driven by technology. More specifically, by the development and universal availability of fast computers and of devices to collect and store ever-increasing amounts of data. Satellite remote sensing, large-scale sensor networks, continuous environmental monitoring, medical imaging, micro-arrays, the various genomes, and computerized surveys have not just created a need for new statistical techniques. These new forms of massive data collection also require efficient implementation of these new techniques in software. Thus development of statistical software has become more and more important in the last decades. Large data sets also create new problems of their own. In the early days, in which the t-test reigned, including the data in a published article was easy, and reproducing the results of the analysis did not take much effort. In fact, it was usually enough to provide the values of a small number of sufficient statistics. This is clearly no longer the case. Large data sets require a great deal of manipulation before they are ready for analysis, and the more complicated data analysis techniques often use special-purpose software and some tuning. This makes reproducibility a very significant problem. There is no science without replication, and the weakest form of replication is that two scientists analyzing the same data should arrive at the same results. It is not possible to give a complete overview of all available statistical software. We started this overview of statistical software (food and nutrition) by indicating that the computer revolution has driven much of the recent development of statistics, by increasing the size and availability of data. Replacement of mainframes by minis, and eventually by powerful personal computers, has determined the directions in the development of statistical software. In more recent times the internet revolution has accelerated these trends, and is changing the way scientific knowledge, of which statistical software is just one example, is disseminated. Due to the development of applied statistics and software engineering, we will get easiest, multifunction operating software in food and nutritional science. Which will make fast our research.

ACKNOWLEDGMENT

The authors thank Dr. Wei Ma, Dr. Kou Xing Ran, Dr. Wang Peng, Zhang Luyue, Qiangwei Wang, Food Nutrition and Functional Factors Research Centre, School of Food Science & Technology, Jiangnan University, Wuxi, P.R. China. And Md. Saifullah, Universiti Putra Malaysia (UPM).

REFERENCES:

Abdel-Aziz H. O., Murai Y., Takasaki I., Tabuchi Y., Zheng H.C., Nomoto K., Takahashi H., Tsuneyama K., Kato I., Hsu D.K., Liu F.T., Hiraga K., Takano Y., 2008 "Targeted disruption of the galectin-3 gene results in decreased susceptibility to NNK-induced lung tumorigenesis: an oligonucleotide microarray study" Journal of Cancer Research and Clinical Oncology 134:777-788.

Aitkin M., Dorothy A., Brian F., John H., 1989 "Statistical Modelling in GLIM." Oxford: Oxford University Press. ISBN 0-19-852203-7.

Argyrous G.,1997 "Statistics for Research: With a Guide to SPSS, SAGE", London, ISBN 1-4129-1948-7

Arthur L., 2013 "Handbook of SAS DATA Step Programming.

CRC Press." pp. 149–165. ISBN 978-1-4665-5238-8

Basak T., Lou Z., Rahman M. R. T., Azam M. S., 2014 "Quality Evaluation of Dehydrated (Sun and Solar Drying) Cabbage and Rehydration Properties, EUROPEAN ACADEMIC RESEARCH ",Vol. II: 6, 8214-8243

Buck D., 2013 "A Hands-On Introduction to SAS DATA Step Programming". SUGI 30: SAS Institute.

Church S., Krines C., 2008 "Using nutritional data in the food industry." Food Sci. Technol., 22(2): 23-25.

Cline M.S., Smoot M., Cerami E., Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T and Bader GD (2007) Integration of biological networks and gene expression data using Cytoscape. Nature Protocols 2:2366-2382.

Committee on toxicity of chemicals in food consumer products and the environment (cot), 2009, pathway analysis software for the interpretation of complex datasets, tox/2009/35

Dennis G, Sherman D.T., Hosack D.A., Yang J., Gao W., Lane H.C., Lempicki R.A., 2003 "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome Biology 4:R60.

Du L., Ke Y., 2013 "Research and Development on Food Nutrition Statistical Analysis Software System". Advance Journal of Food Science and Technology 5(12): 1637-1640, 2013 ISSN: 2042-4868; e-ISSN: 2042-4876

Encyclopedia of Research Design, 2010, doi:10.4135/9781412961288. ISBN 9781412961271

Francis, B., Mick G., Clive P., 1993, "The GLIM System: Release 4 Manual." Oxford: Clarendon Press. ISBN 0-19-852231-2.

Fukushima T., Kikkawa R., Hamada Y., Horii I. 2006, "Genomic cluster and network analysis for predictive screening for hepatotoxicity". Journal of Toxicological Sciences 31:419-432.

Gerecke D.R., Chen M., Isukapalli S.S., Gordon M.K., Chang Y.C., Tong W., Androulakis I.P., Georgopoulos P.G., 2009, "Differential gene expression profiling of mouse skin after sulfur mustard exposure: Extended time response and inhibitor effect." Toxicol Appl Pharmacol 234:156-165.

Gunawardana C.G., Kuk C., Smith C.R., Batruch I., Soosaipillai A., Diamandis E.P., 2009, "Comprehensive analysis of conditioned media from ovarian cancer cell lines identifies novel candidate markers of epithelial ovarian cancer." J Proteome Res 8:4705-4713.

Hardcastle W.A., 2000, "A Guide to the Configuration and Customisation of Software used in Laboratories," Publisher : LGC

Huang D.W., Sherman B.T., Lempicki R.A., 2009 "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources". Nature Protocols 4:44-57.

Jan D., Lee W., 2009, "Statistical Software – Overview"

John A.B., 1995, "Statistics for food science I," Nutrition & Food Science 4: 19–23 © MCB University Press · ISSN 0034-6659

Jothi, J. S., Hashem S., Rana M. R., Rahman M. R. T. Shams-Ud-Din, M. 2014, "Effect of Gluten-free Composite Flour on Physico-chemical and Sensory Properties of Cracker Biscuits," J. Sci. Res. 6 (3), 521-530 (2014)

Koroušic S.B., Stibilj V., Pograjc L., Mis N.F., Benedik E., 2013, "Food composition databases for effective quality nutritional care." Food Chem., 140(3): 553-561.

Lambert C.B., Spire C., Renaud M.P., Claude N., Guillouzo A., 2009, "Reproducible chemical-induced changes in gene expression profiles in human hepatoma HepaRG cells under various experimental conditions." Toxicology In Vitro 23:466-475.

Levesque R., 2007, "SPSS Programming and Data Management: A Guide for SPSS and SAS Users", SPSS Inc., Chicago Ill. ISBN 1-56827-390-8

Lora D.D, Susan J.S., 2012, "The Little SAS Book: A Primer : a Programming Approach." SAS Institute. p. 6. ISBN 978-1-61290-400-9

Margolis D., Land J.W.H., Gottlieb R., and Qiao X. 2011. "A complex adaptive system using statistical learning theory as an inline preprocess for clinical survival analysis." Procedia Comput. Sci., 6: 279-284.

Mauritsen, R.H., 1984, Logistic regression with random effects. Ph.D. thesis, Department of Biostatistics, University of Washington.

Min Y., 2001, "A review of random effects models in EGRET for Windows (Version 2.0.3)", Centre for Multilevel Modelling, Institute of Education, University of London m.yang@ioe.ac.uk, egret.doc

Naito Y., Kuroda M., Mizushima K., Takagi T., Handa O., Kokura S., Yoshida N., Ichikawa H., Yoshikawa T., 2007, "Transcriptome analysis for cytoprotective actions of rebamipide against indomethacin-induced gastric mucosal injury in rats." Journal of Clinical Biochemistry and Nutrition 41:202-210.

Nelder J., 1975, "Announcement by the Working Party on Statistical Computing: GLIM (Generalized Linear Interactive Modelling Program)". *Journal of the Royal Statistical Society, Series C* 24 (2): 259–261. JSTOR 2346575.

Ouyang B., 2013. "Theory of regression apple professional cooperation organization research." Adv. J. Food Sci. Technol., 5(3): 328-331.

Peilong X, Na N., Qing L., 2013. Effects of low-ester pectin on wheat dough's physical properties detected by rheological method. Adv. J. Food Sci. Technol., 5(8): 1011-1014.

Pereira, A.C., Marco S.R., Pedro M., 2009, "Quality control of food products using image analysis and multivariate statistical tools." Ind. Eng. Chem. Res., 48(2): 988-998.

Spector P.,1995, "An Introduction to the SAS System"; Statistical Computing Facility; Department of Statistics; University of California, Berkeley

Rahman M. R. T., Tang Y., Wang Q. and Al-Hajj N. Q. M., 2014, Short Review : Statistics and Different Departments of Food Industry, IJCBS REVIEW PAPER, 1 :3, 41-47

Rocke D.M., Ideker T., Troyanskaya O., Quackenbush J., Dopazo J., Papers on normalization, variable selection, classification or clustering of microarray data. Bioinformatics 25:701-702.

Royal Statistical Society webpage on Working Parties at the Wayback Machine(archived February 21, 2007)

SAS, 2010, Institute Inc. and World Programming Limited *(England and Wales High Court (Chancery Division)*

Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T., 2003, "Cytoscape: A software environment for integrated models of biomolecular interaction networks". Genome Research 13:2498-2504.

Sharma A.K., Searfoss G.H., Reams R.Y., Jordan W.H., Snyder P.W., Chiang A.Y., Jolly R.A., Ryan T.P., 2009, "Kainic acid-induced F-344 rat model of mesial temporal lobe epilepsy: gene expression and canonical pathways". Toxicol Pathol 37:776-789.

SPSS, 2006, 15.0 Command Syntax Reference 2006, SPSS Inc., Chicago Ill.

Vladislav V., Gitesh R., Michael M., Andrew M., Carla B., Steven P., 2011 "Development of mobile-accessible nutritional system to improve healthy food choices." Res. J. Appl. Sci. Eng. Technol., 3(9): 1034-1047.

Wacholder S., 1986, "Binomial regression in GLIM: Estimating risk ratios and risk differences". American Journal of Epidemiology 123 (1): 174–184.PMID 3509965.

Wolstenholme D., Obrien C., Nelder J., 1988, "GLIMPSE: a knowledge-based front end for statistical analysis". Knowledge-Based Systems 1 (3): 173. doi:10.1016/0950-7051(88)90075-5

Youssef Z., 2012, "Introduction to STATA, Data Statistics Analysi ,University Academic Computing Technologies" ; The American University in Cairo