

Statistika a řízení jakosti s využitím WWW

*Jiří Militký a Aleš Linka**

Motto:

Na Webu je všechno (optimista).

Na Webu není nic (pesimista).

Na Webu je moc užitečného (laik).

Na Webu jsou známé věci (specialista).

Co je tedy na Webu??

- INFORMACE (a informace o informacích)
- SOFTWARE (Shareware, Freeware, Evaluační verze a software na vyzkoušení)
- HYPERTEXTOVÉ UČEBNICE
- ON-LINE VÝUKA
- VÝPOČETNÍ PROSTŘEDÍ (Applety atd.)
- ČASOPISY
- VIRTUÁLNÍ PRODEJNY

Typická ukázka uspořádaných informací o statistice

MIKE TALBOT

<http://www.bioss.sari.ac.uk/~mike/webtra.htm#qual>

STATISTICAL TRAINING ON THE WEB

Lists of WWW lists tam jsou spojení dále po WWW.

Training on-line

INTRODUCTORY

Methodology : Mathematics, General introductory, Descriptive statistics, Sampling,

Probability and inference, Regression, Multivariate analysis,

Design of

experiments, Time series analysis, Quality Control, Non-parametric, Bayesian.

Applications : Business management, Psychology, Education, Medicine, Engineering, Biology, Physical Sciences.

ADVANCED

Graphics

Software on-line : Java applets, other programs.

Software training : SAS, Minitab, S-Plus, R, Genstat, SPSS, Excel, Bugs.

Software - general :

Books & Journals

Glossaries

Data sources

Case studies

History

* KTM, Textilní fakulta TU Liberec, jiri.militky@vslib.cz, ales.linka@vslib.cz

Organisations

Training initiatives

General educational matters

METHODOLOGY - GENERAL INTRODUCTORY

- HyperStat Online : D. Lane, Rice University
- *Study of Stability in Variation* : J. de Leeuw, UCLA
- SurfStat Australia : K. Dear, University of Newcastle
- Introductory Statistics : D. Stockburger, Southwest Missouri SU
- Introductory Statistics Demonstrations : Ohio University
- Berrie's MultiMedia Statistics Page:
- Statistical tests : University of Amsterdam
- Resources for Teachers of Probability and Statistics : University of Illinois.

A mnoho dalších. Řada hypertextových příruček je propojena na software (např. XLISP) nebo JAVA Applety

APPLICATION - MEDICINE

- Introduction to Biomedical Statistics : T. Rossini, Univ. of Wisconsin.
- Stats in preventive medicine : M. Janairo, Loyola University Medical Center

APPLICATION - PHYSICAL SCIENCES

- Remote Sensing and Data Analysis : STAR, Asian Institute of Technology, Thailand

ADVANCED TRAINING

- Data Analysis BriefBook : R.Bock & W.Krischer, CERN, Geneva
- G.A.S.P. - Globally Accessible Statistical Procedures: S. Carolina
- **SMART : a statistical explorapaedia** : BioSS, UK
- Resampling Stats: Univ. of Maryland
- Data visualization and image collection : R.Purman, NHGSST
- Teaching Wavelets in XploRe : S. Klink,
- Neural Network FAQ: W. Sarle, SAS
- Introduction to Monte Carlo Methods : Comp. Sci. Educ. Proj.
- The Bartlett correction page : F. Cribari-Neto, S. Illinois Univ.
- Bayesian Knowledge Discovery Project : P.Sebastiani, City University
- Introduction to Wavelets : Amara Graps, Max-Planck-Institut fur Kernphysik

SOFTWARE ON-LINE

Software On-Line - Java, Javascript et al

- Links to lots of software on the Web : J.C. Pezzullo
- Interesting Java Applets : Duke Univ.
- Java applets for teaching : W. West, Univ. of South Carolina
- Illustrating statistical concepts with Java : B. Zielman, NL
- Mathematics Archives with Java : University of Kentucky
- Statistics Applications with Java : University of Kentucky
- Exploring regression and other techniques : J.Puranen, Helsinki Univ.
- **Statlets : Statistical analysis and graphics** : NWP Associates
- Demo of design optimisation algorithm : MultiSimplex KB

SOFTWARE ON-LINE - MISCELLANEOUS TOOLS

- Statistical tests : University of Amsterdam
- Links to software providers : Duke University
- Pexlab - planning and design of experiments : University of Leeds
- Non-parametric statistical testing : Univ. of Victoria
- MANET- graphical exploration of data : Univ. of Augsburg

Software - SAS

Software - Minitab

Software - S-Plus

Software - R

Software - Genstat

Software - SPSS

Software - Excel

Evaluation of survey packages : ASA Survey

DATA SOURCES

- CMU StatLib (USA) : Carnegie Mellon's archives
- Statistical Resources on the Web : Univ. of Michigan
- Data Surfing on the World-Wide Web : R. Lock, St Lawrence Univ.
- Data for course in statistics : Chatterjee, Handcock and Simonoff,
- Finding data on the Internet : Robert Niles

HISTORY

- MendelWeb: classical genetics, data analysis, plant science.
- History of Mathematics : St Andrews University, Scotland
- Some Famous Statisticians : J. de Leeuw, UCLA

GENERAL EDUCATIONAL RESOURCES

- WWW Virtual Library - Educational Technology : Univ. of Geneva
- WWW Virtual Library - Education : Charles Sturt Univ.
- Resources for Teachers : Robert Gordon Univ. Scotland

V dalším jsou uvedeny typické resp. z našeho hlediska zajímavé příklady z jednotlivých oblastí (to co je označeno šedě) a něco navíc. Jsou zde demonstrovány typy výstupů tak jak je poskytuje WWW bez úprav a překladu. Pouze některé podrobné informace jsou vynechány. Očekáváme, že zvědavý čtenář se zřejmě napojí na uvedené adresy a získá to, co ho zajímá přímo ze zdroje. Na konci textu je zmíněn projekt, kterým se budeme zabývat v příštím roce. Jsou vítány všechny připomínky kontakty a názory.

Učebnice a příručky

<http://www.stat.ucla.edu/textbook/>

Statistics The Study of Stability in Variation

Page author: Jan de Leeuw [deleeuw@stat.ucla.edu]

I. Introduction

II. Analysis of a Single Variable

III. Analysis of a Pair of Variables

IV. Analysis of Multivariables

You need Xlisp-Stat locally



Příklad výkladu jednoho pojmu:

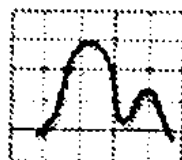
The Mid-Range

The mid-range is the value in the middle of the range, i.e. it is equally far from the minimum as it is from the maximum observation. In a formula, we have

$$M = (x_{\max} + x_{\min})/2,$$

The mid-range only depends on the largest and smallest values, it does not care about the values in between Text pokračuje asi na dvě stránky. V řadě případů jsou k dispozici ilustrativní příklady v Xlisp-Stat. Tato učebnice se nám jeví jako velmi zajímavá a psaná na slušné úrovni.

Typický a špičkový příklad
firemní příručky jako doprovod
k programovému balíku:



ELECTRONIC TEXTBOOK
StatSoft

<http://www.statsoft.com/textbook/stathome.html>

- DEMO 1: The Interactive STATISTICA Application with the Electronic Manual
- DEMO 2: The STATISTICA Guided Tour (recorded movie)
- DEMO 3: The SEWSS Movie (interactive movie)

DEMO 1: The Interactive STATISTICA Application with the Electronic Manual:
This program is a complete application (and not a self-running demo) with many features of Basic Statistics module of STATISTICA, and it allows you to run your own analyses and produce customized graphs.

Downloadable Electronic Statistics Textbook

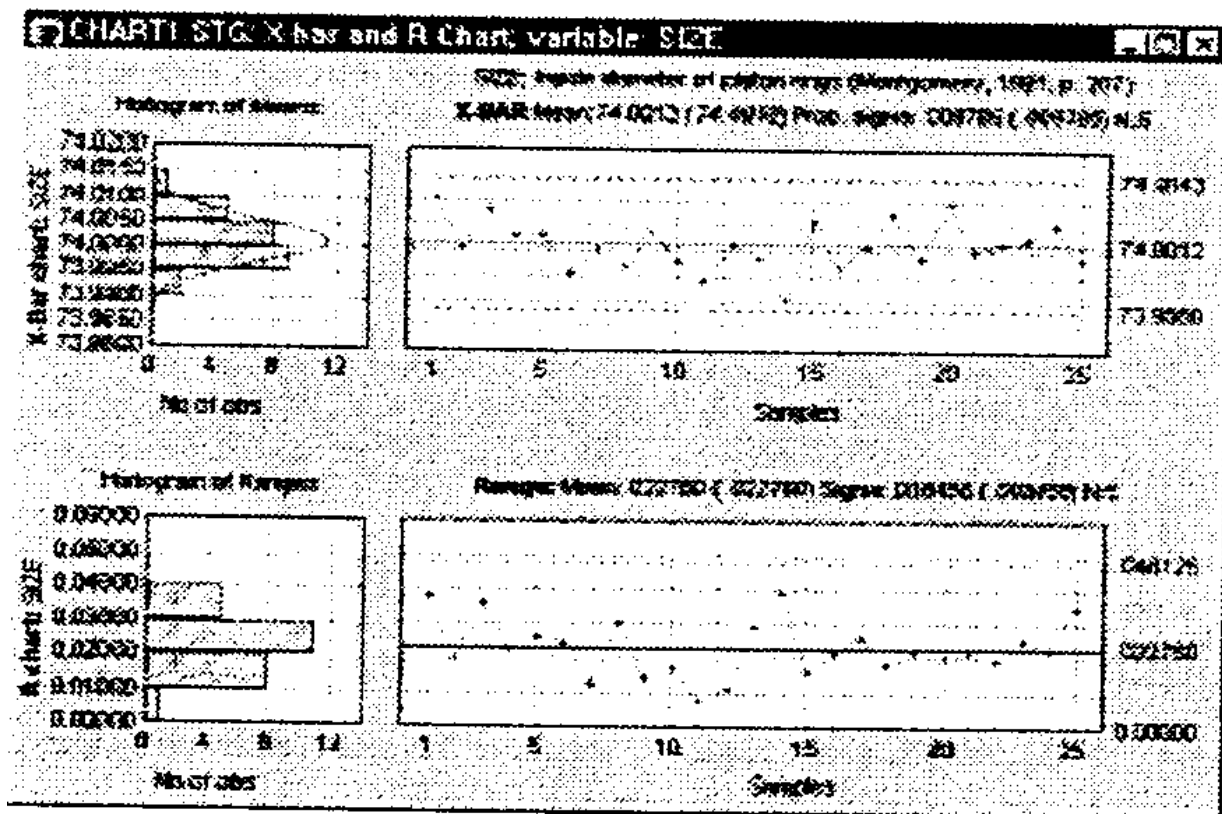
This Electronic Statistical Textbook offers training in the understanding and application of statistics. The material was developed at the StatSoft R&D department based on many years of teaching undergraduate and graduate statistics courses and covers a wide variety of applications, including laboratory research (biomedical, agricultural, etc.), business statistics and forecasting, social science statistics and survey research, data mining, engineering and quality control applications, and many others.

Typický výstup jedné kapitoly:

- Quality Control Charts
- General Purpose
- General Approach
- Establishing Control Limits
- Common Types of Charts
- Short Run Control Charts
 - Short Run Charts for Variables
 - Short Run Charts for Attributes
- Multiple Stream Group Charts
- Unequal Sample Sizes
- Control Charts for Variables vs. Charts for Attributes
- Control Charts for Individual Observations
- Out-of-Control Process: Runs Tests
- Operating Characteristic (OC) Curves
- Process Capability Indices
- Other Specialized Control Charts

General Purpose

In all production processes, we need to monitor the extent to which our products meet specifications. In the most general terms, there are two "enemies" of product quality: (1) deviations from target specifications, and (2) excessive variability around target specifications. The most standard display actually contains two charts (and two histograms); one is called an X-bar chart, the other is called an R chart. Pokračování je na několik stránek. Jsou k dispozici i ilustrativní výstupy programu STATISTICA



Applety v jazyku JAVA

Statlets Home Page obsahuje soubor Java appletů . Datové soubory do 100 řádků a 10 sloupců lze zpracovat zdarma. Jinak je třeba se přihlásit a platit.

Distributions

- Histograms
- Normal Approximation to Binomial
- Normal Density
- The T Distribution
- Area Under Normal Curves
- Z Scores & the Normal Distribution

Probability & Stochastic Processes

- Binomial Probabilities
- Brownian Motion
- Central Limit Theorem
- A Gamma Process
- Let's Make a Deal Game

Statistics

- Guide to basic stats labs
- ANOVA
- Confidence Intervals
- Regression
- Spearman's rank correlation
- Statistics Tools
- Statoscope Summarizing data
- T-test

Models

- Triogram Models

Examples

- Epidemic Study
- Survival Analysis Example
- An Ising Model

Other

- Interactive Quadratic Surface rendering
- Simple Least-Squares Regression
- Psychological Statistics (scatterplots, regression, discriminant analysis)

Příklad: Regression Applet

The applet below is designed to teach students the effect of leverage points on a regression line. Students may add points to the plot by clicking the mouse button. Students should note that adding points close to the existing line barely changes the line. By adding points far from the existing line, the regression line changes considerably. This is particularly true for points added outside the range of the data. This should help students understand the effect of outliers on regression analysis

Komplexní statistický systém pracující přes WWW prohlížeče

WebStat : <http://www.stat.sc.edu/~west/webstat/version2.0beta/>

Statistický programový systém napsaný jako Java applet umožňující statistickou analýzu na World Wide Web. Je třeba pouze WWW prohlížeč schopný práce s Java applety.

Na horní liště tohoto okna je nabídka obsahující 5 položek:

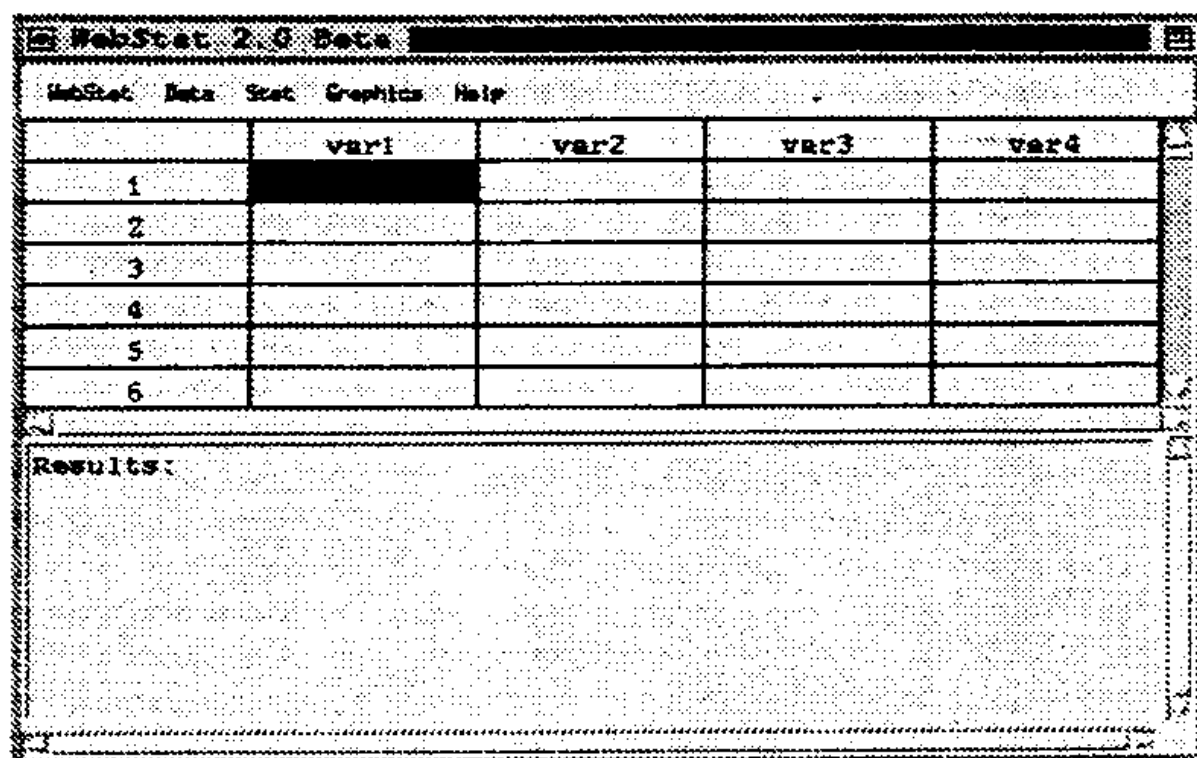
WebStat - general information about the software and the option to exit the software.

Data - Methods for loading data, accessing sample data sets, clearing data, obtaining more information about data and saving data.

Stat - Access to all of the specific data analysis procedures the software provides as well as the options for saving/printing results from these procedures.

Graphics - Access to all of the specific graphical procedures the software provides and the option for showing the Graphics Window.

Help - the option to display software documentation.



The screenshot shows a window titled "WebStat 2.0 Data". It has a menu bar with "WebStat", "Data", "Stat", "Graphics", and "Help". Below the menu bar is a table with 5 columns: "vari", "var2", "var3", and "var4". The first column is numbered 1 to 6. The table contains the following data:

	vari	var2	var3	var4
1				
2				
3				
4				
5				
6				

Below the table is a section labeled "Results:" which is currently empty.

Atraktivní alternativa (pro nespecialisty v JAVA aplikacích)

MATLAB WEB server 1.0

<http://www.mathworks.com/products/webserver/highlights.shtml>

Pro rozšíření aplikací MATLABu přes WWW. S využitím standardních HTML dokumentů umožňuje tento produkt propojit programy v MATLABu s WWW. Dokumenty HTML slouží jako grafické uživatelské rozhraní (GUI) pro aplikace v MATLABu. Uživatel nepotřebuje znalost jazyka MATLAB a tento jazyk není ani na jeho počítači instalován.

Požadavky

MATLAB 5.3

server s operačním systémem Windows NT, resp. Sun Solaris

software **MATLAB WEB server 1.0**

Základní rysy

- nástroj pro zpracování informací v HTML (čtení dat přímo z HTML, vkládání výsledků včetně grafických do HTML formulářů přes Web prohlížeč)
- Výpočty na serverech - aplikace MATLAB je spuštěn a pouze na speciálním serveru a uživatel jej nemá na svém počítači instalován
- Rozhraní MATLABu přes HTML, - připravené pomocí vhodného editoru

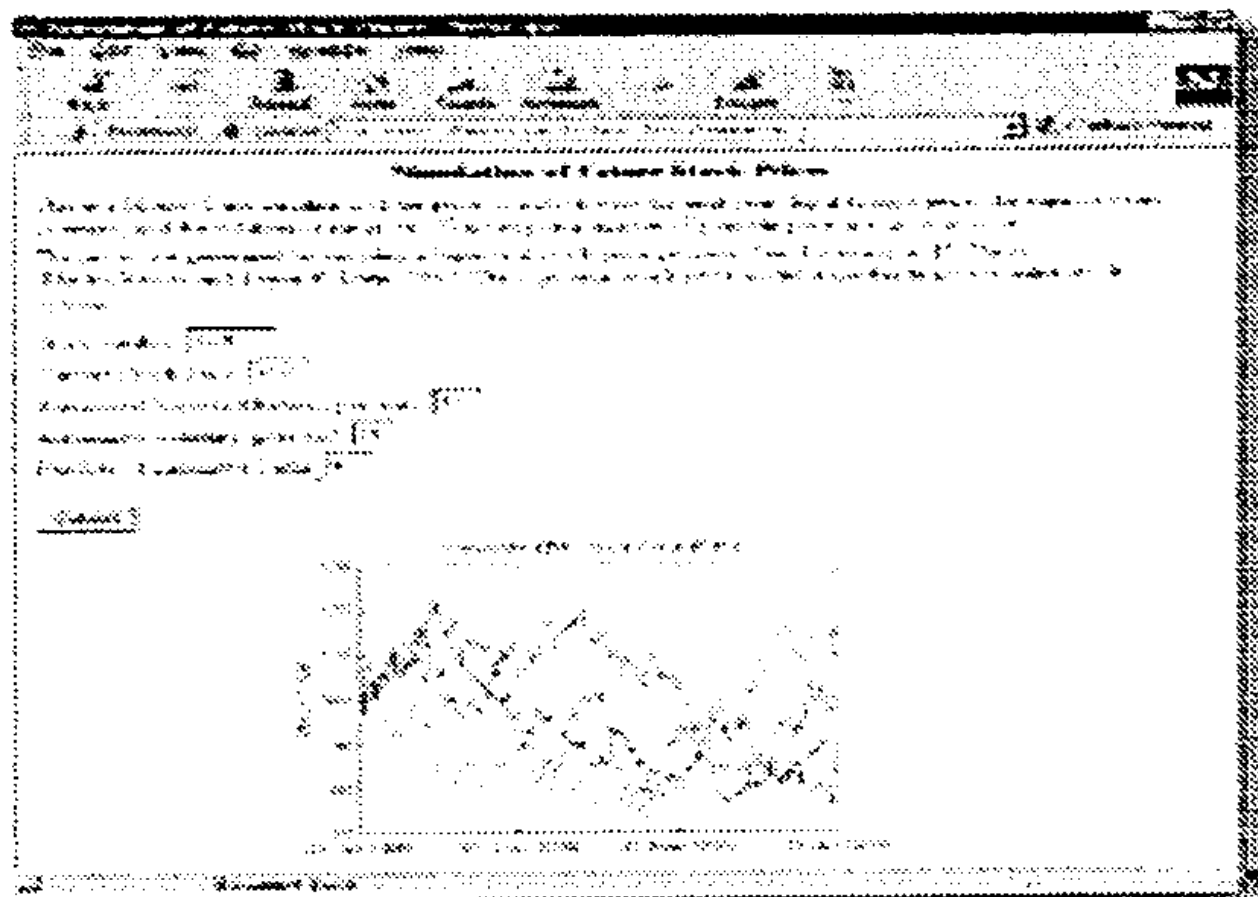
(Microsoft

FrontPage).

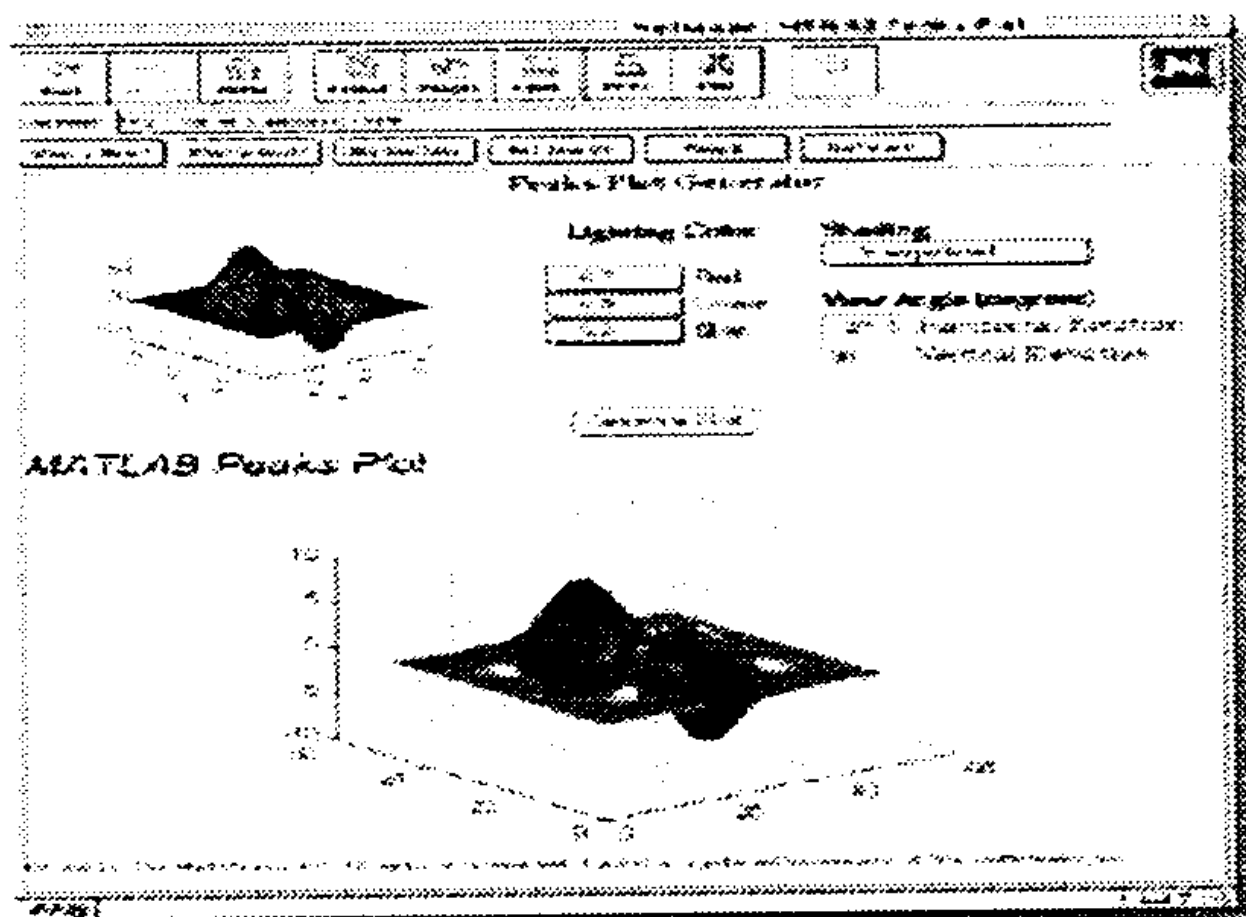
- Použití pokročilého jazyka pro generaci tabulek a grafů

Příklady výstupů :

Tool for Monte Carlo simulation of a stock price process being accessed by a common Web browser.



MATLAB function peaks return a sophisticated 3-D plot.



Projekt SMART

www.bioss.sari.ac.uk/smart/unix/monthome.htm

(Statistics and Mathematics as Advanced Research Tools) slouží k seznámení specialistů s počítačově orientovanými statistickými metodami prostřednictvím WWW. Využívá se multimediální presentace. Určeno především pro biology a biotechnology

Znaky systému SMART:

- umožnění přípravy odborníků ne statistiků v daném čase
- nabídka technik 10-20 min přehled s možností hlubšího studia
- orientace na aplikace tj. vysvětlování na problémech známých odborníkům v dané oblasti
- spojení s vhodným software pro podporu výpočtů
- vysoká modularita - tj. adaptace pro různé situace
- propojení statistiků se specialisty v oboru pomocí WWW
- použití běžného software a hardware (standardní prohlížeče)
- cenová efektivita s ohledem na zdroje a přípravu výukových materiálů

Struktura systému:

- software pro interakci s uživatelem (JAVA)
- HTML texty grafika a audio soubory
- protokoly pro přípravu, distribuci a přístup k prezentacím

Pouze malá část je v jazyce JAVA

Moduly mají tři úrovniovou strukturu

1. základní úroveň série obrázků (každý na jednu obrazovku). Slouží k rychlému přehledu o technikách obsažených v daném modulu. Motivační příklady. Popis cílů, výsledků a interpretace
2. Detailní objasnění složitých částí a organizace větvením s možností návratu do první úrovně.
3. Reference a poznámky přístupné z libovolné stránky

Typická stránka

- rychlý souhrn v textové formě
- vazby na stránky s detaily
- návaznost na stránky, kde jsou připomínky uživatelů
- návaznost na ilustrace a obrázky

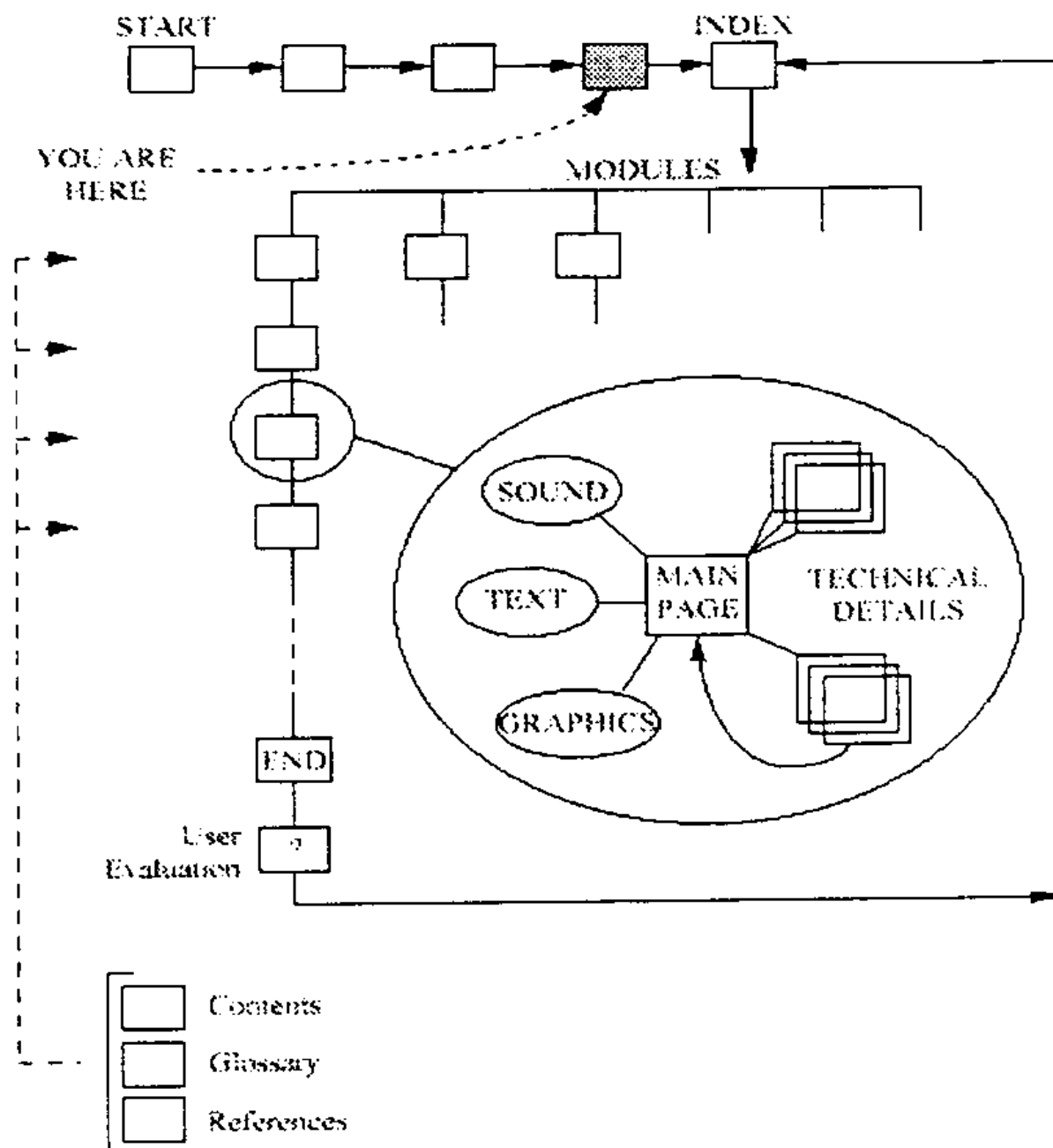
Návaznost na software

1. vstup dat spuštění programu a získání výsledků s využitím webového prohlížeče
2. zadání dat a příkazů přes webový prohlížeč přímo do programů, které jsou na uivatelové počítači (SAS, S Plus, XLISP)

SMART

Explorapedia of Statistical and Mathematical Techniques for use in Research and Technology

Systém propojení jednotlivých částí



Moduly SMART

Growth curve modelling
Antedependence (repeated measures) modelling
Analysis of molecular variance
Sequential acceptance sampling
Variogram analysis
Partial least squares regression
Generalized additive models
Finding over/under-represented words in DNA
Creating your own module
New : Introduction to Basic Statistics Using Genstat / Excel

Methodologies

Survey design and analysis, experimental design and analysis, linear and non-linear modelling, multivariate methods, non-parametric methods, time series, spatial statistics, image analysis.

Příklad GROWTH CURVE MODELLING

Growth is a fundamental property of biological systems, occurring at the level of populations, individual animals and plants, as well as within organisms. Much research has been devoted to modelling growth processes, and there are many ways of doing this, including: mechanistic models, time series, stochastic differential equations etc.

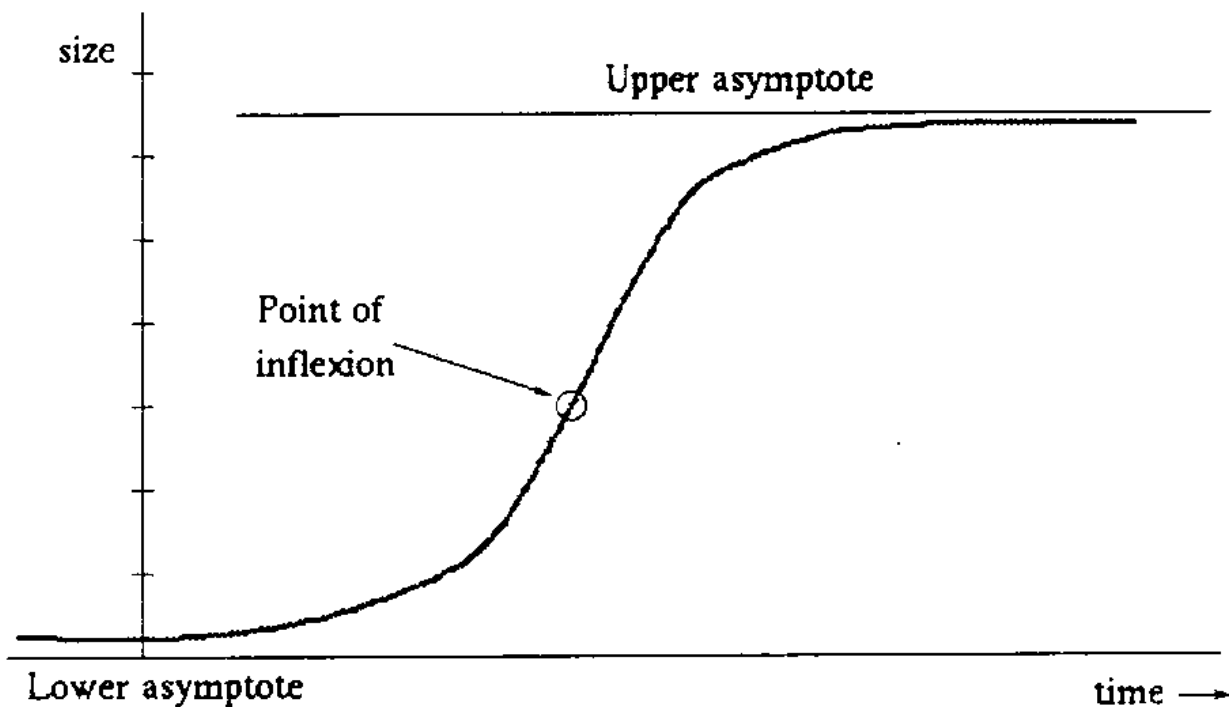
Examples of applications of the technique include:

animal weights over time;

growth of human populations;

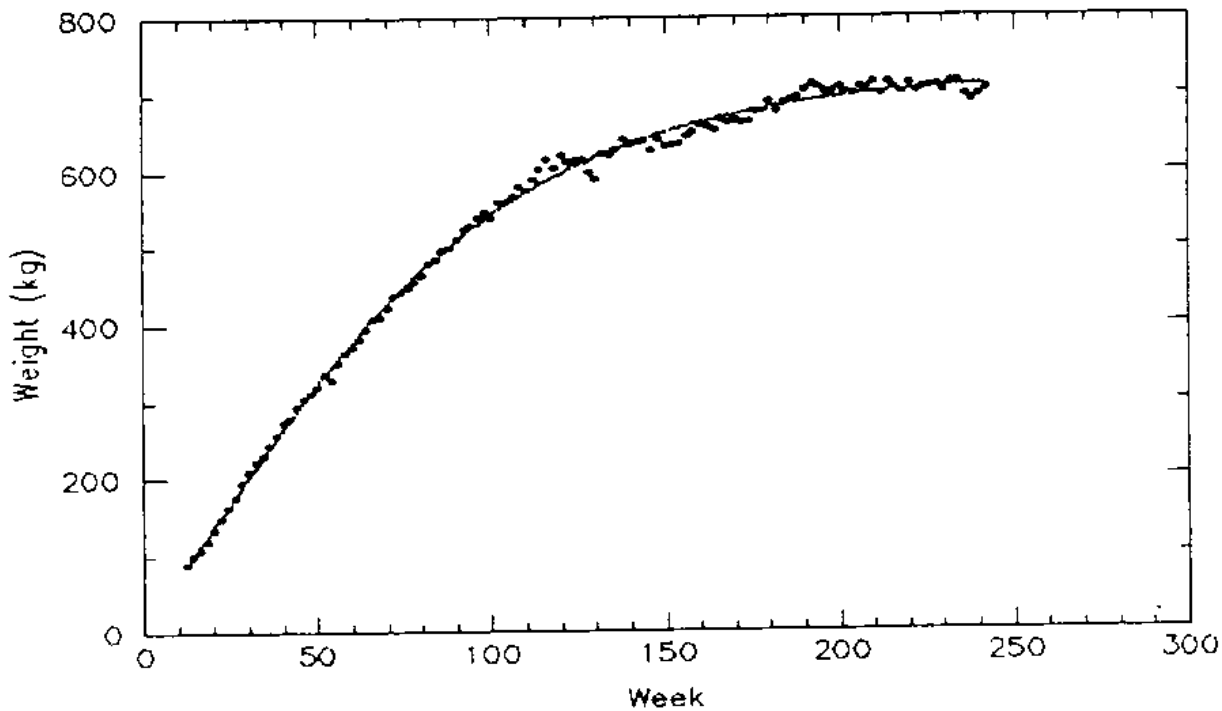
biomedical data;

development of organisms.



APPLICATION TO COW WEIGHTS

We will fit the generalised logistic curve to body weight data for each cow. Cow live weights have been recorded every two weeks for 232 weeks, i.e. 116 observations in total.



This is the output from analysis of a cow's weight data using the Genstat 5 package.

The formula for the generalised logistic takes the form :

$$Y = A + \frac{C}{\left(1 + Te^{-B(x-M)}\right)^{1/T}}$$

where Y = weight, height, size, etc... and X = time

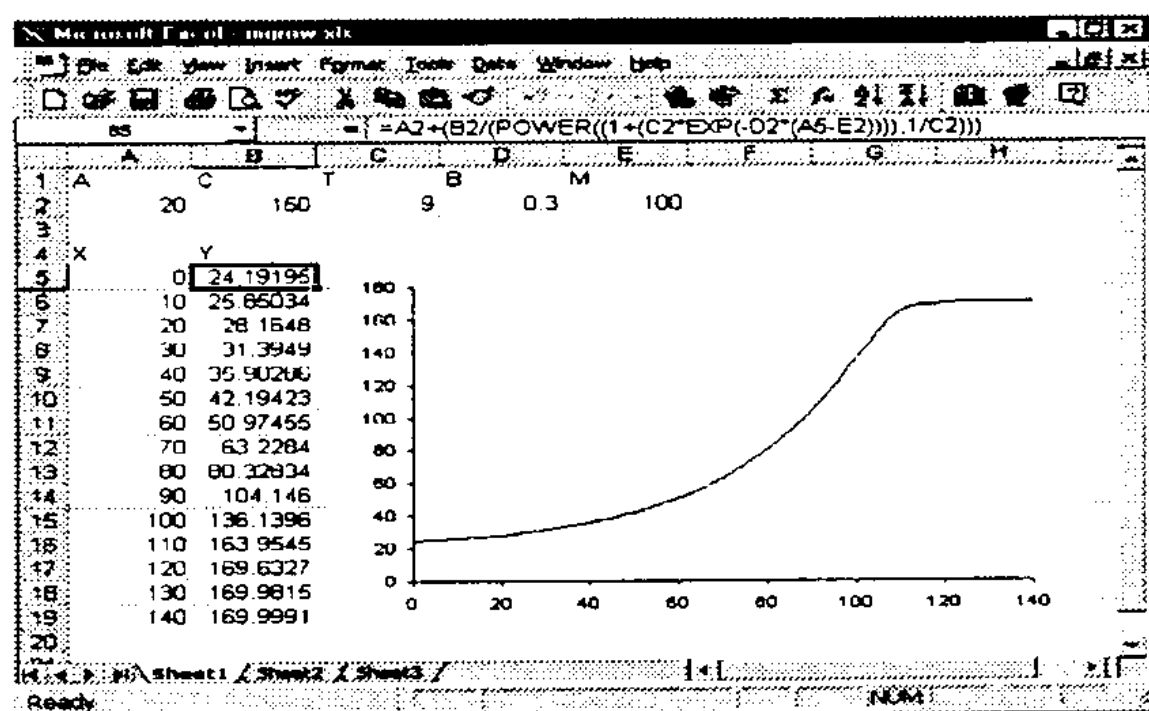
* The parameters have the following roles:

- A - controls the lower asymptote;
- C - controls the upper asymptote;
- M - controls the time of maximum growth;
- B - controls the growth rate;
- T - controls where the maximum growth occurs.

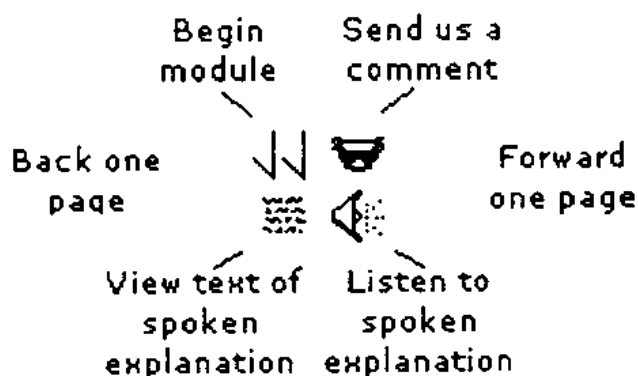
Základní nabídka

- GROWTH CURVE MODELLING
- INTRODUCTION
- Growth curve modelling
- Example: Animal weights
 - A typical growth curve
- GENERALISED LOGISTIC
 - The generalised logistic curve
 - Application to cow weights
 - Fitting generalised logistic
 - Fitted curve
 - Residuals from generalised logistic
- GOMPERTZ
 - to samé jako u předchoziho
- How to Choose a Model ?
- Associated links
- CONCLUSION

Instrukce pro výpočet v EXCELU a výstup



Ovládání SMART:



Náš plán:

Příprava hypertextové učebnice **Nástroje pro řízení jakosti** o řízení jakosti s využitím jazyka **MATLAB**. Obsah bude pravděpodobně tento:

- Základní statistika (exploratorní, robustní, adaptivní)
- Ověření předpokladů o datech
- ANOVA
- Autokorelace a trendy
- Transformace dat
- Simulace dat s předepsaným rozdělením
- Regulační diagramy

Očekáváme využití software **MATLAB WEB server 1.0** a zkušenosti autorů software **QC Expert**

QC.Expert -typický výstup

