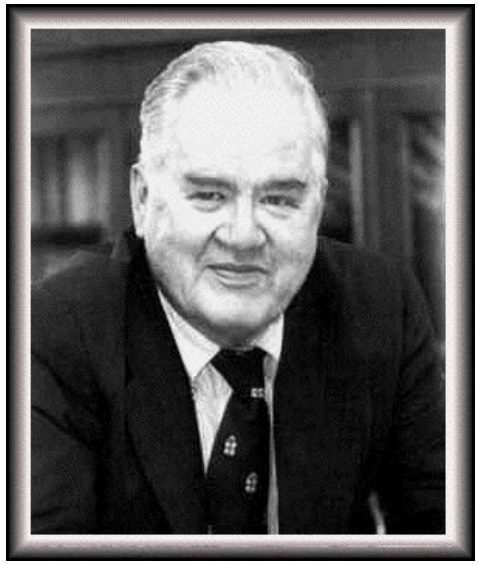
Integrable *e*-lements for Statistics Education

Wolfgang Härdle
Sigbert Klinke
Uwe Ziegenhagen

Institute für Statistics and Econometrics
Humboldt-Universität zu Berlin
http://ise.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

"Each new generation of computers offers us new possibilities, at a time when we are far from using most of the possibilities offered by those already obsolete."

John W. Tukey (1965)



e-lements in Statistics Education

Modern education in statistics must involve practical computer-based data analysis.

Questions

- Which elements do re-occur during different courses?
- Which technology can be presented during class, which not?
- High-level code at the beginning of the studies?
- Where are the limits of *e*-lements in statistics education?

Outline

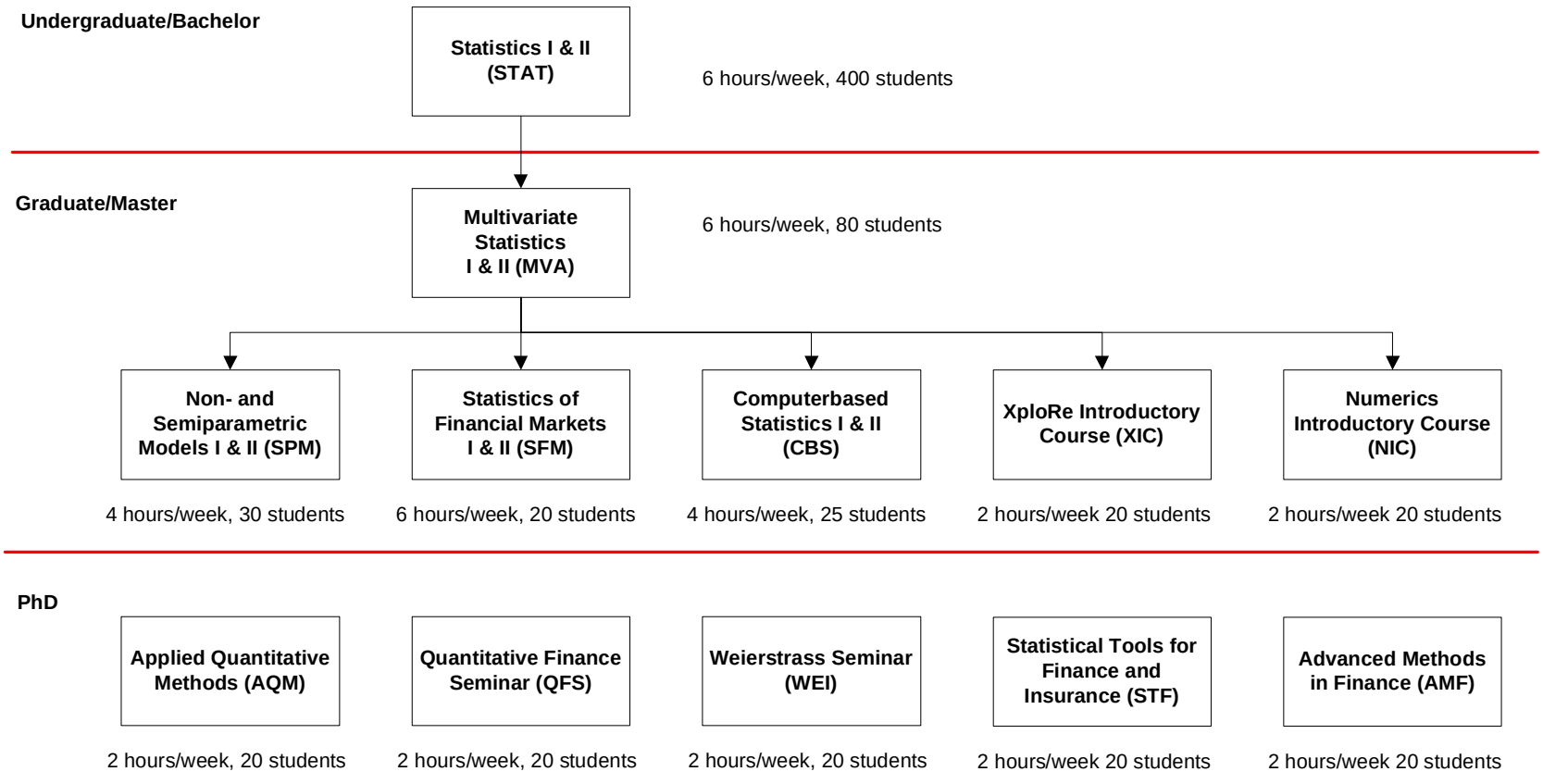
- Introduction ✓
- Statistics courses
- MM*Stat and e-stat
- Electronic books
- XploRe and Yxilon
- Limits of *e*-lements

CASE Courses

Students with different backgrounds are taught at ISE:

- German Business Administration and Economics
- BA/MA courses in Economics and Statistics
- Students from math and other science departments

Layout of Studies



Statistics I & II

- Basic probability theory
- Random variables
- Discrete and continous distributions
- Point and interval estimation
- OLS-Regression
- Analysis of timeseries

Multivariate Statistics I & II

- Histograms and kernel density estimation
- Matrix algebra and multivariate distribution
- Principal component and discriminant analysis
- Cluster analysis and multidimensional scaling
- Factor analysis and projection pursuit

Statistics of Financial Markets I & II

- Options and derivatives
- Black-Scholes model
- Exotic options
- Financial time series
- Value at Risk and copulae

Applied Quantitative Methods

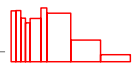
- Analysis and interpretation of multivariate data
- Generalized linear models
- Statistical process control and trend detection
- Computer intrusion detection by statistical means
- Architecture of internet search engines

PDF Slides for Undergraduate Studies

Residual Sum of Squares (RSS)

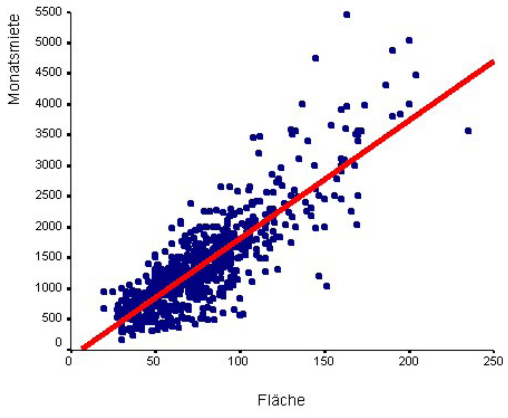
$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to \min \quad | \; \hat{y}_i = b_0 + b_1 x_i$$

$$\begin{aligned} S(b_0, b_1) &= \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \to \min_{b_0, b_1} \\ \frac{\partial S(b_0, b_1)}{\partial b_0} &= -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) \doteq 0 \\ \frac{\partial S(b_0, b_1)}{\partial b_1} &= -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)x_i \doteq 0 \end{aligned}$$



PDF Slides for Undergraduate Studies

lineare Regressionsfunktion: $\hat{y}_i = b_0 + b_1 x_i$



- 815 Berlin flats
- X: m^2
- Y: 1m rent

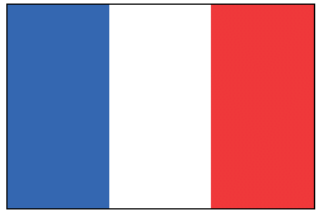
MM*Stat

- support studies at undergraduate level
- HTML-based 'filing cards'
- embedded JavaScript and Java applets
- published by Springer and MHSG (http://www.mhsg.de)

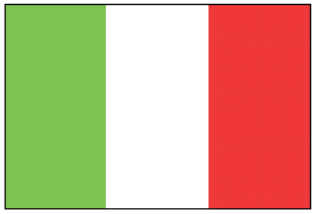
MM*Stat Translations

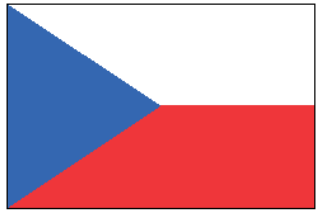

German


English


French

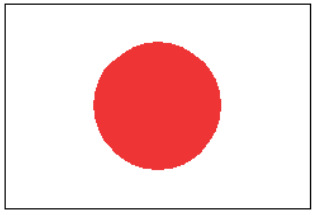

Spanish

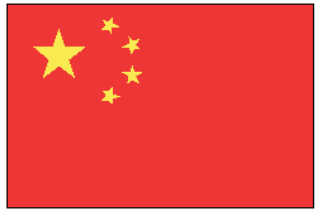

Italian

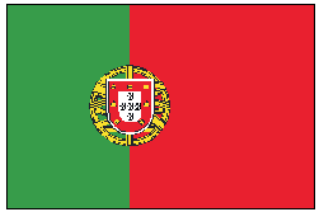

Czech

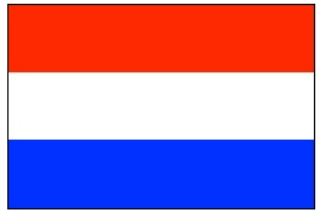

Polish


Indonesian

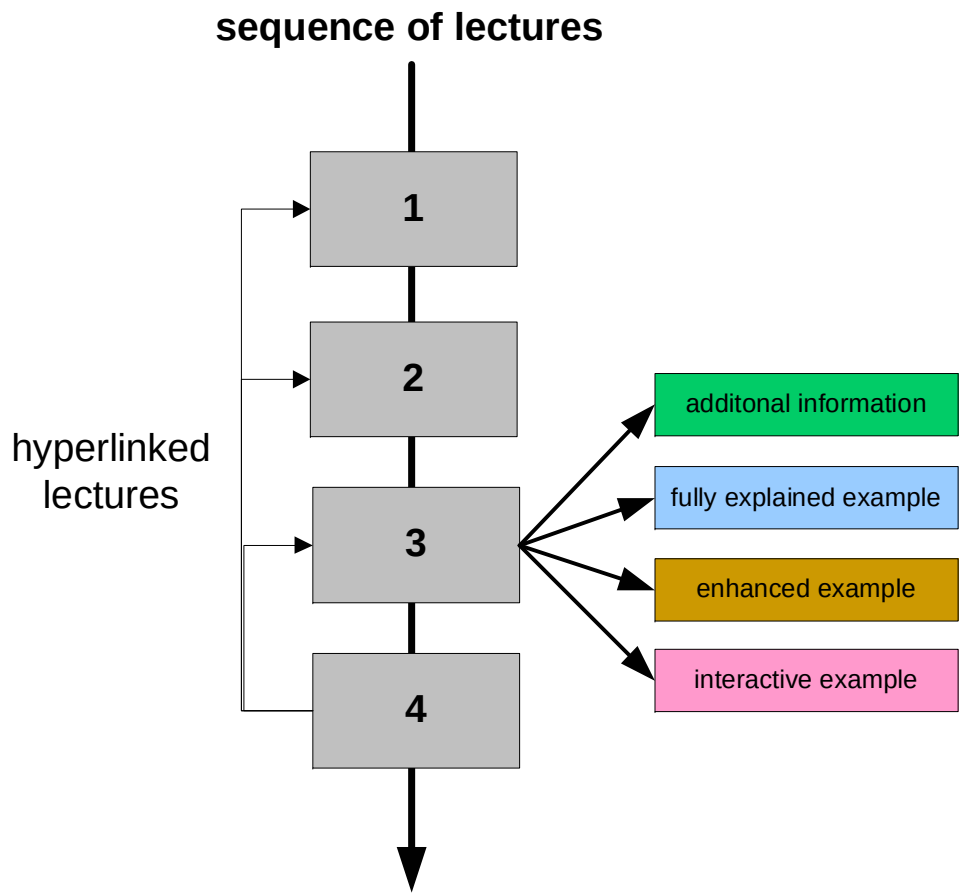

Japanese


Chinese

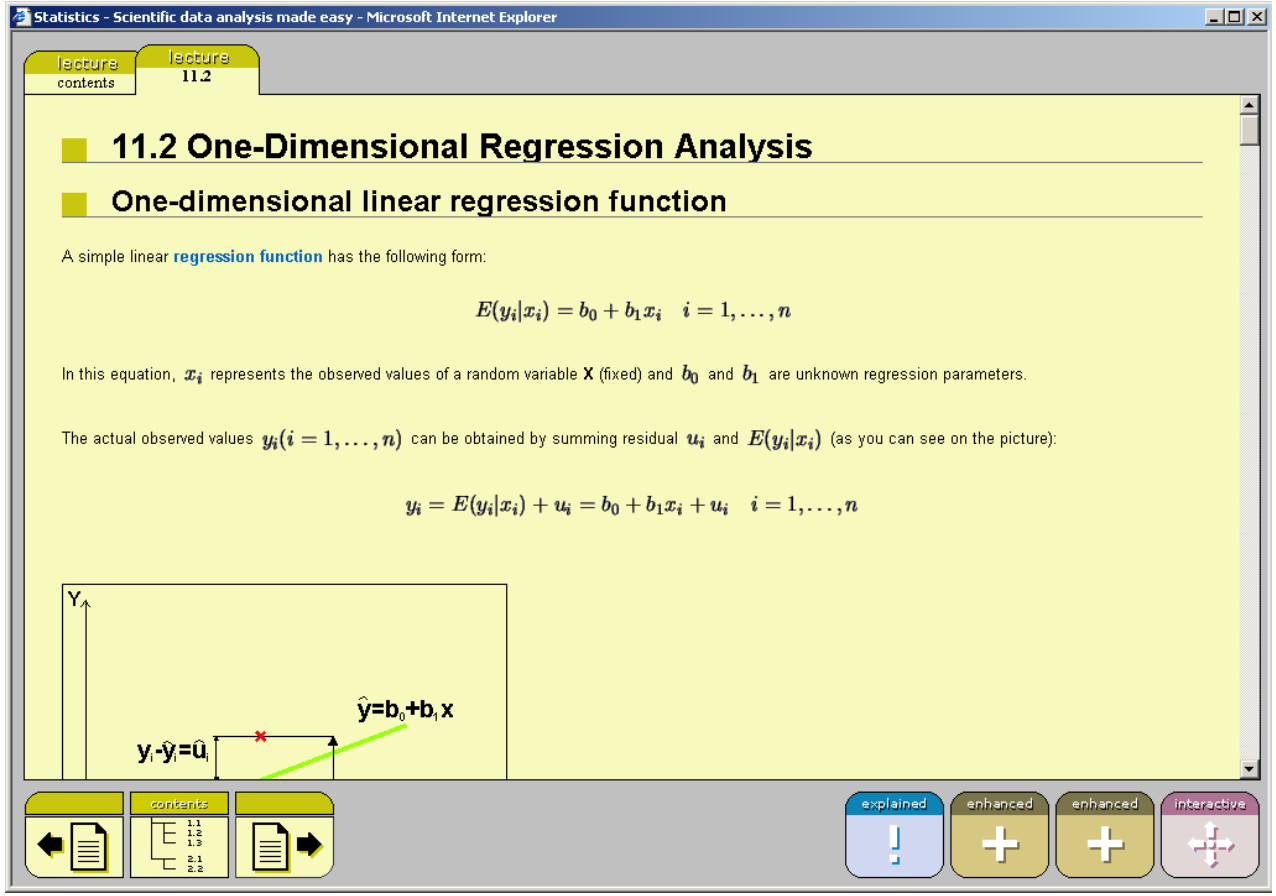

Portugese


Dutch

Different Layers of Learning



MM*Stat Lecture



MM*Stat Explained

Statistics - Scientific data analysis made easy - Microsoft Internet Explorer

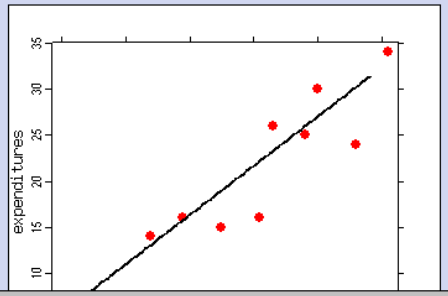
lecture contents | lecture 11.2 | explained 11.2

Example for one-dimensional linear regression

Now, we examine the monthly net income and monthly expenditures on living of 10 two-person households.

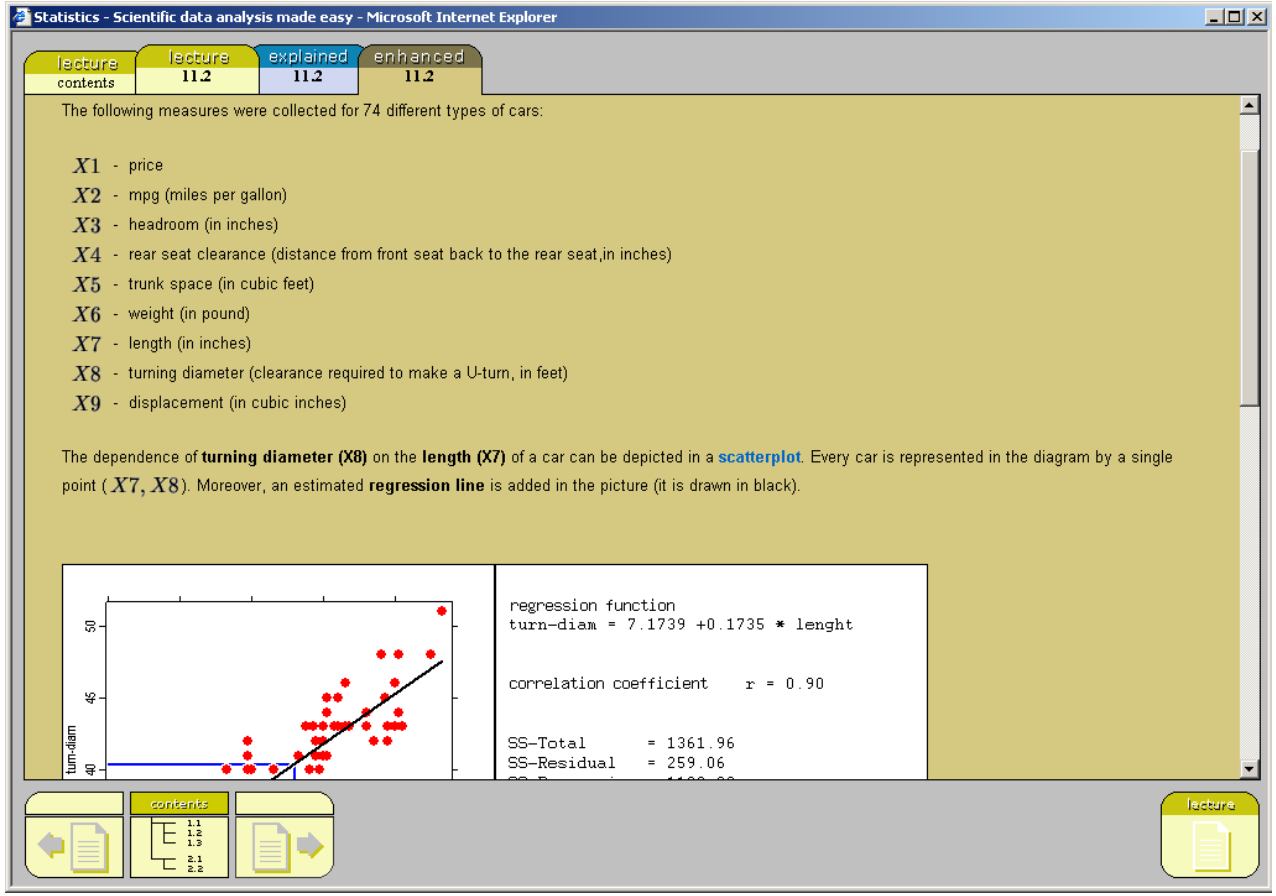
Household	1	2	3	4	5	6	7	8	9	10
Net income in DM x_i	3,500	5,000	4,300	6,100	1,000	4,800	2,900	2,400	5,600	4,100
Expenditures in DM y_i	2,000	3,500	3,100	3,900	900	3,000	2,100	1,900	2,900	2,100

These observations are drawn in the following scatterplot. You can see that the net income of a household has a positive influence of the household's expenditures and that this dependence can be estimated by means of a linear regression function.

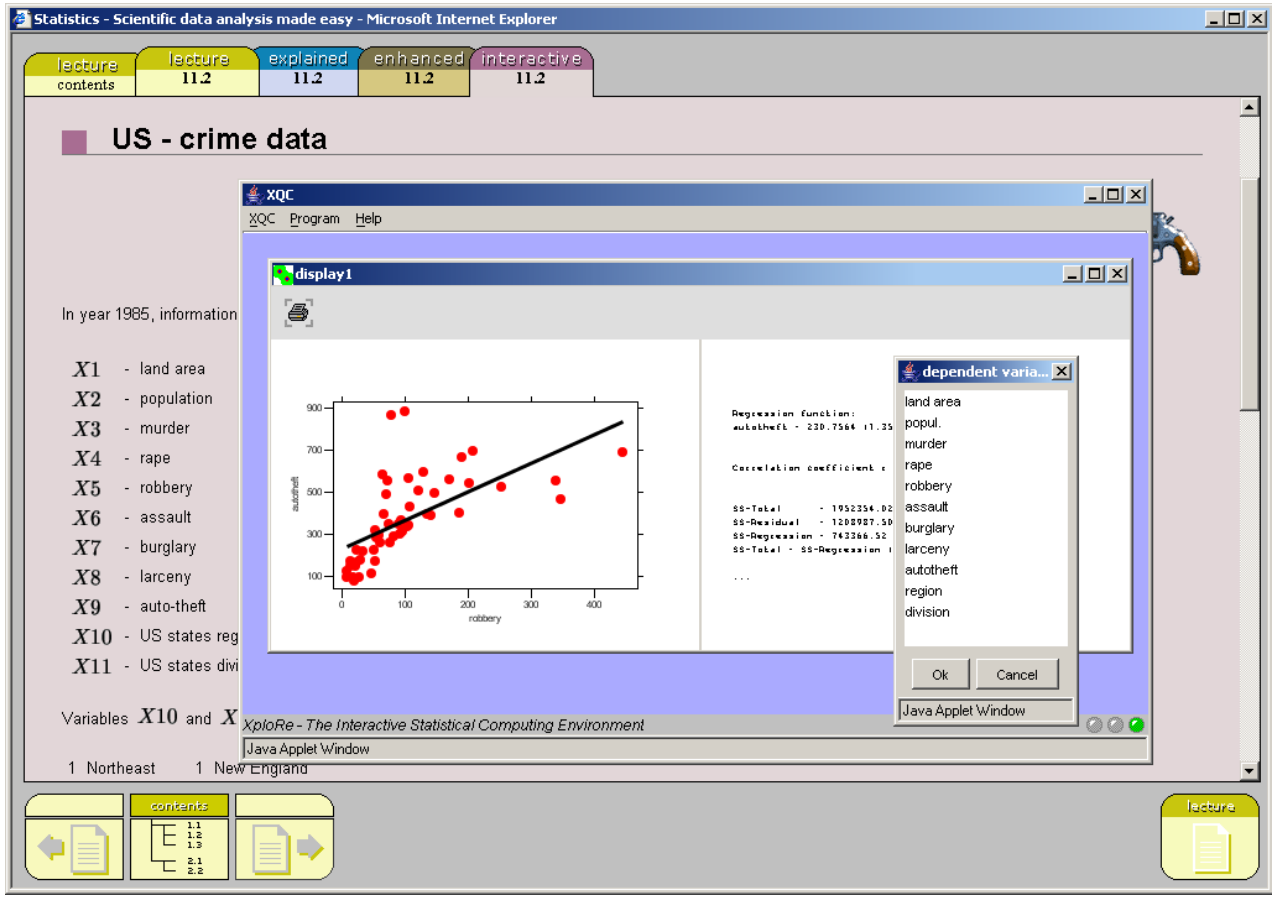


contents | lecture

MM*Stat Enhanced



MM*Stat Interactive

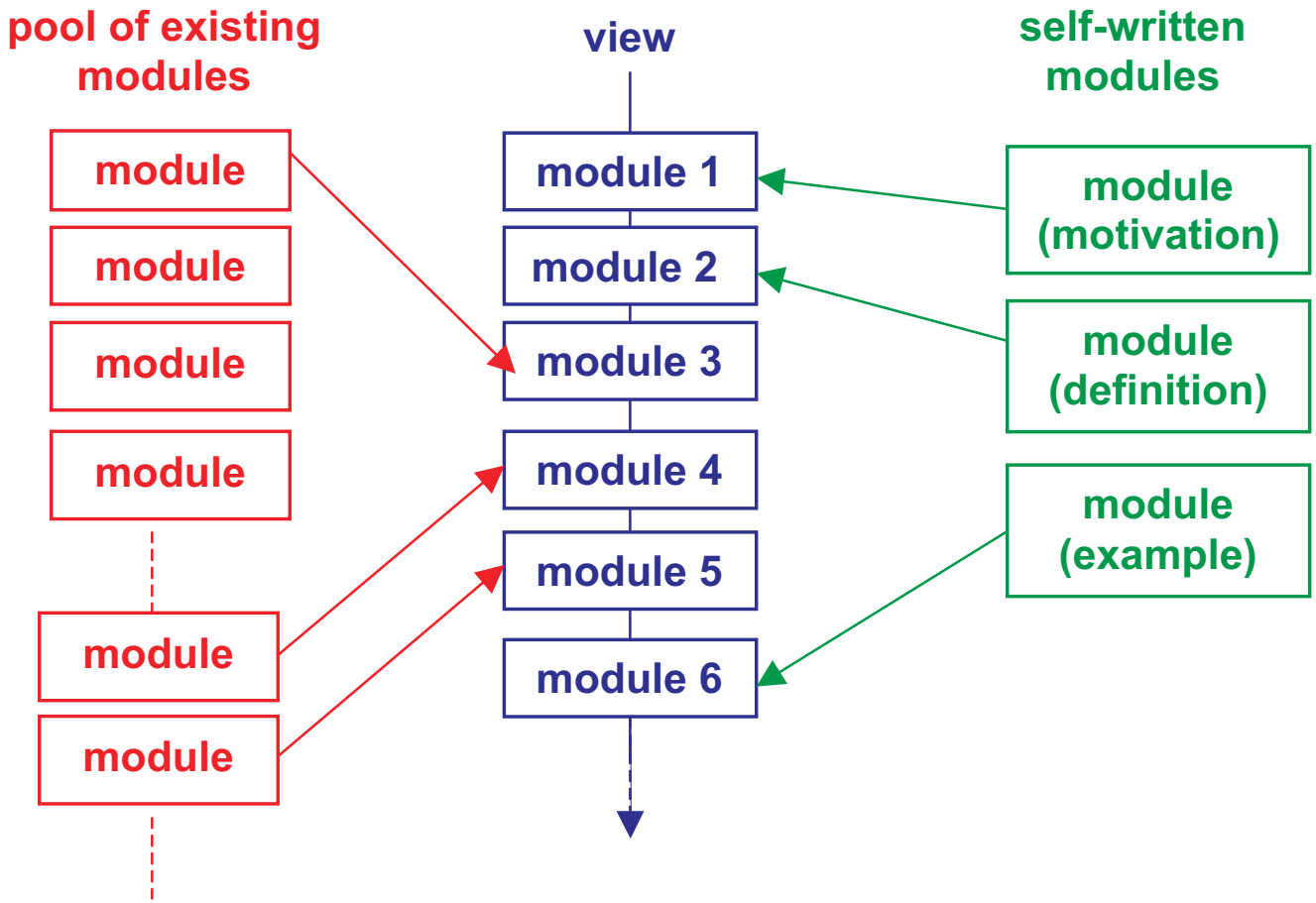


e-stat

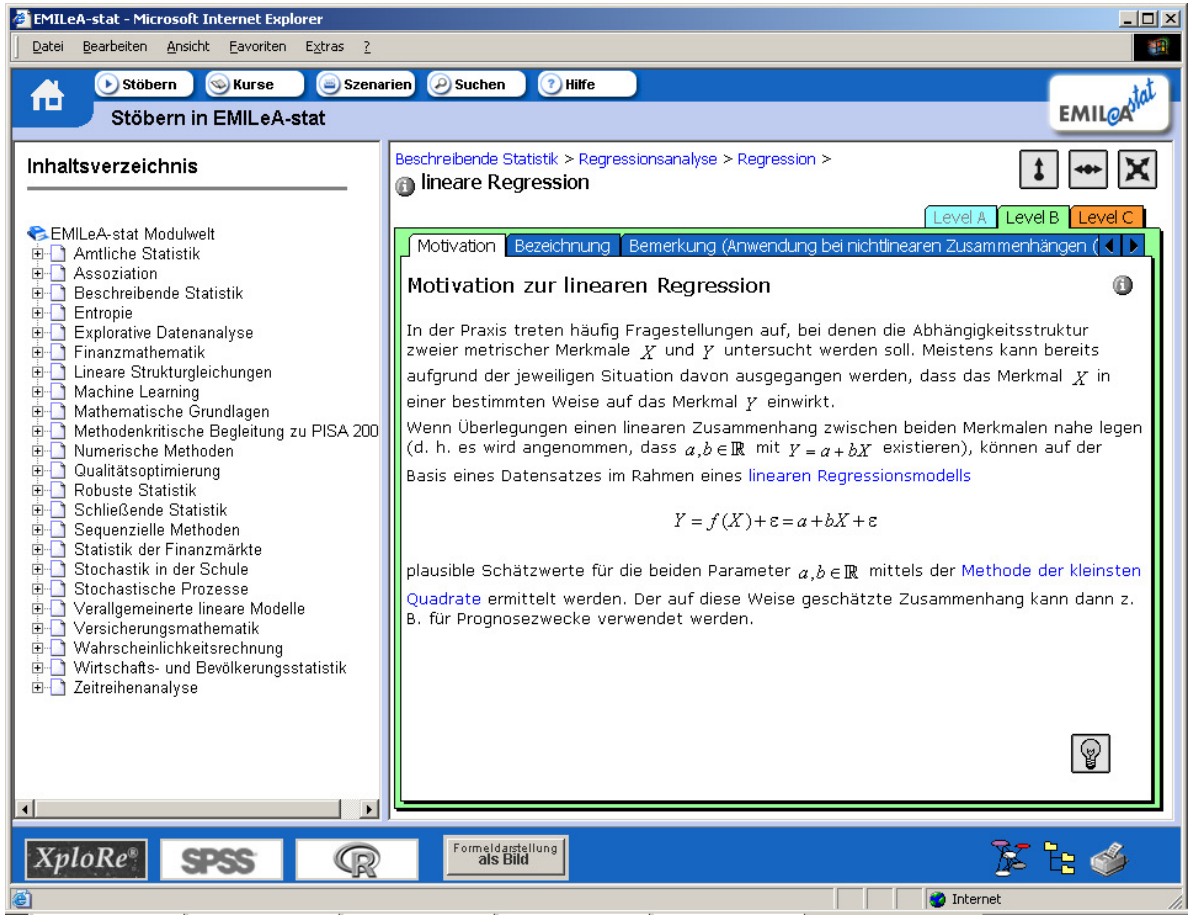
- Developed by team of seven German universities
- Funded by BMBF
- XML-based
- Statistical content broken into small modules
- Example: regression analysis
  1. actual motivation
  2. explanation of general purpose
  3. specification of regression model
  4. listing of properties
  5. estimation techniques

e.stat

e-stat structure



e-stat example

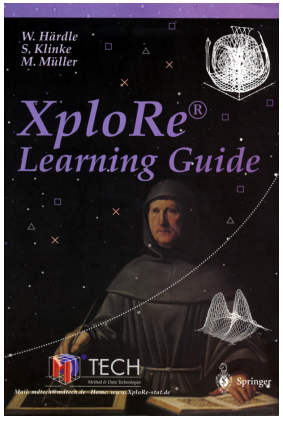


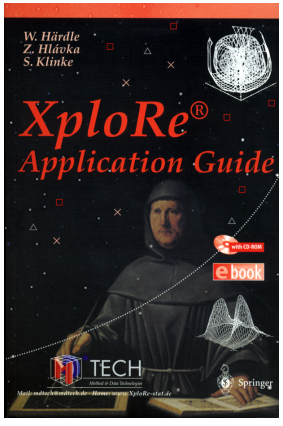
MD*Book Architecture

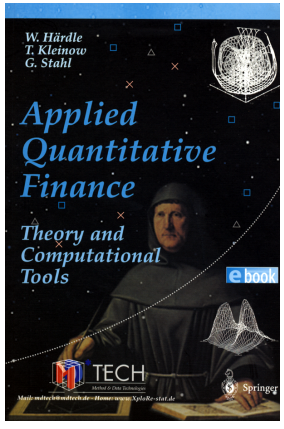
Aim:

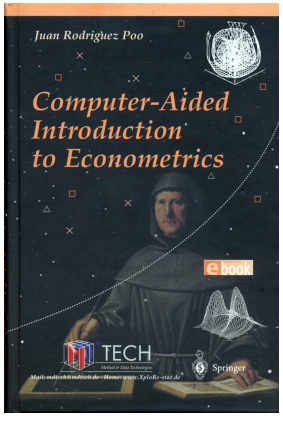
- Generate different formats from one LaTeX source
- PDF, PS, HTML, MD*booklet
- Add interactive examples, visualizing the theory
- for printed books download versions available (incl. XploRe Client/Server)

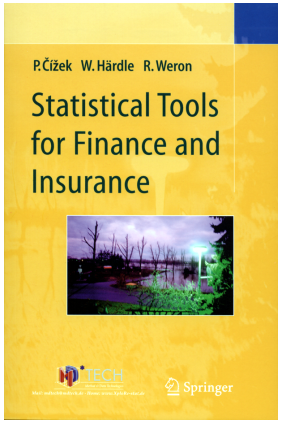
Printed and Electronic Books

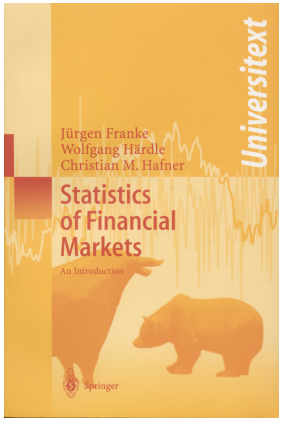


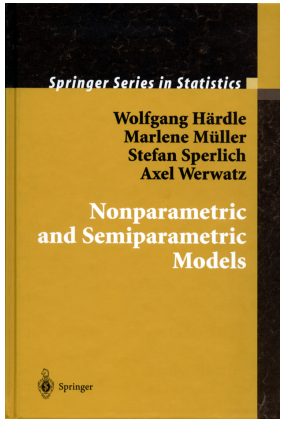


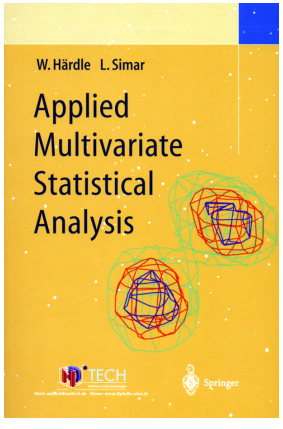




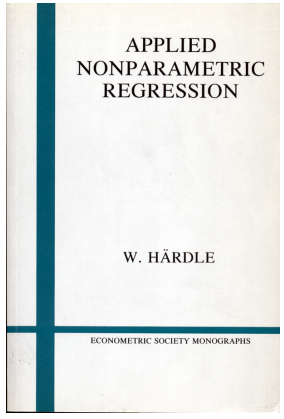


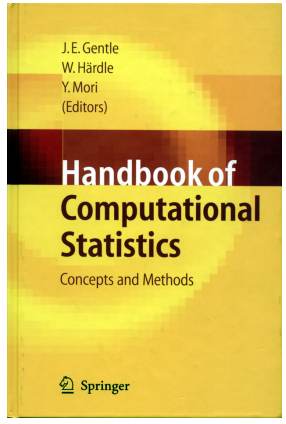


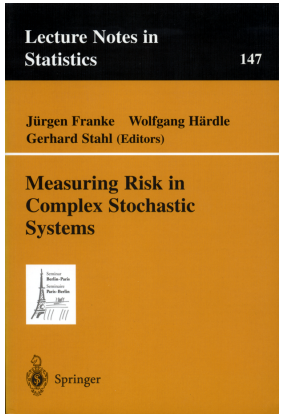


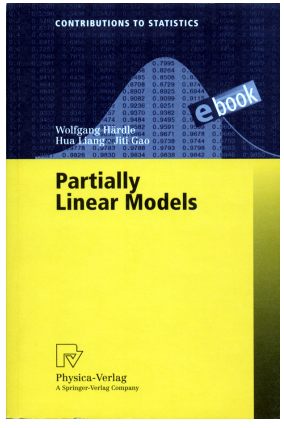


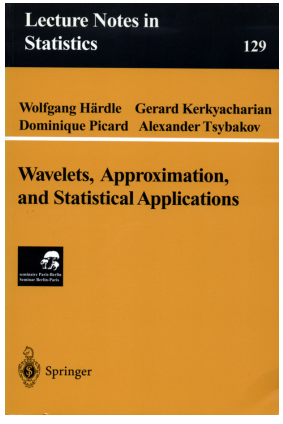
Printed and Electronic Books











Applied Multivariate Analysis - Slide

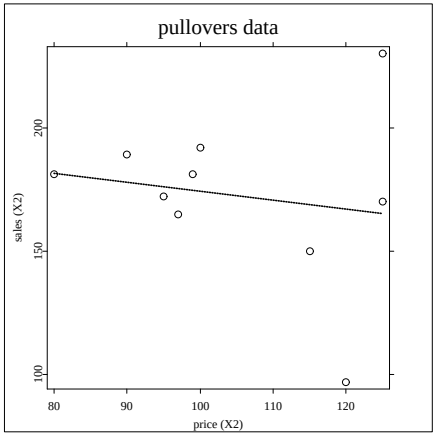
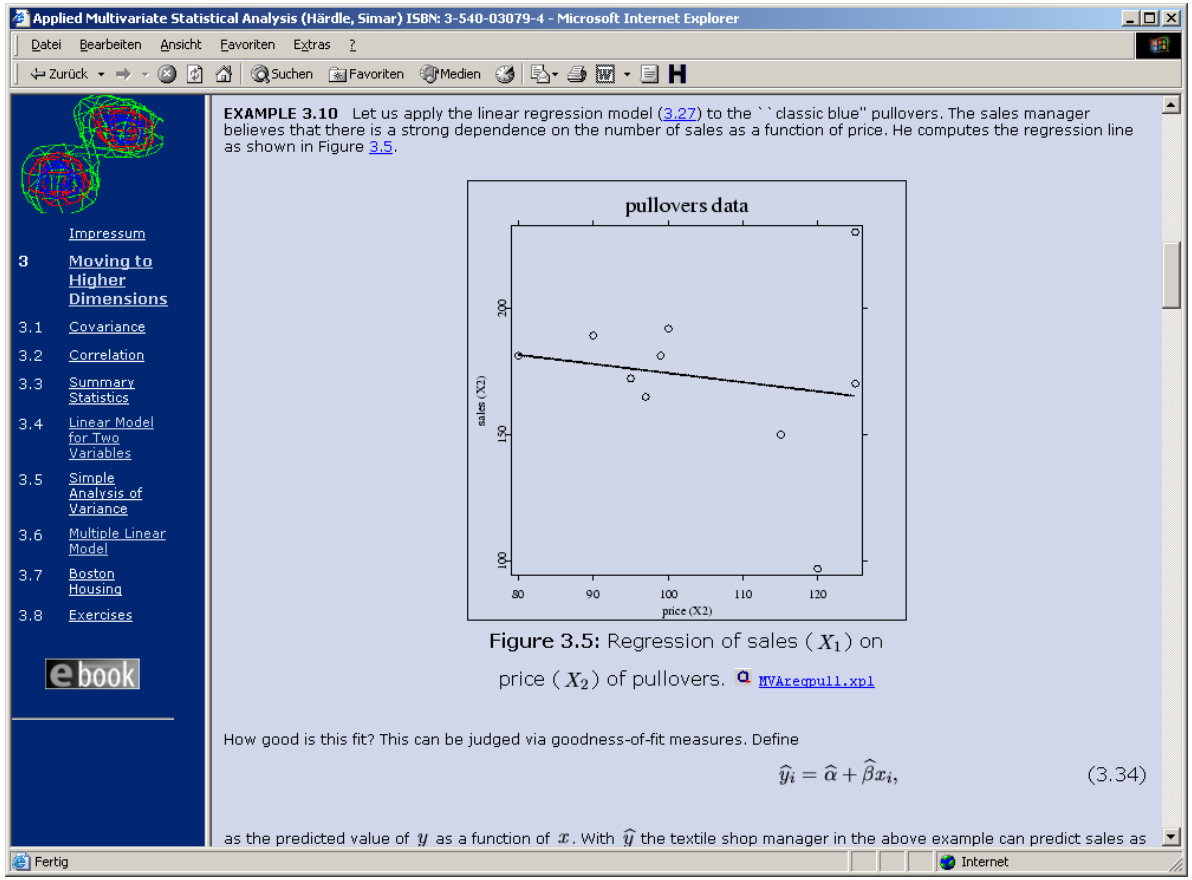
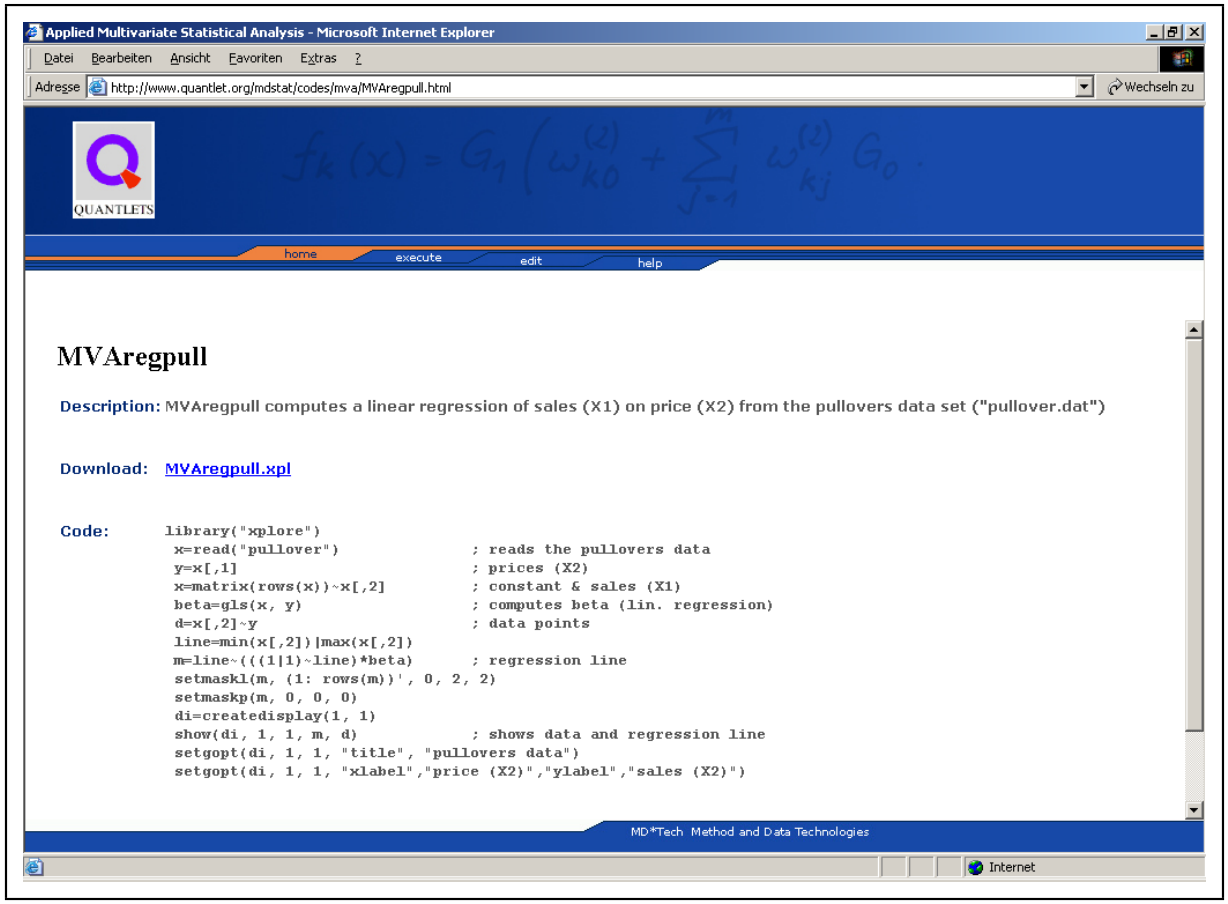


Figure 34. Regression of sales (X_1) on price (X_2) of pullovers, $\widehat{\beta}_0 = 210.7$, $\widehat{\beta}_1 = -0.36$. MVAregpull.xpl

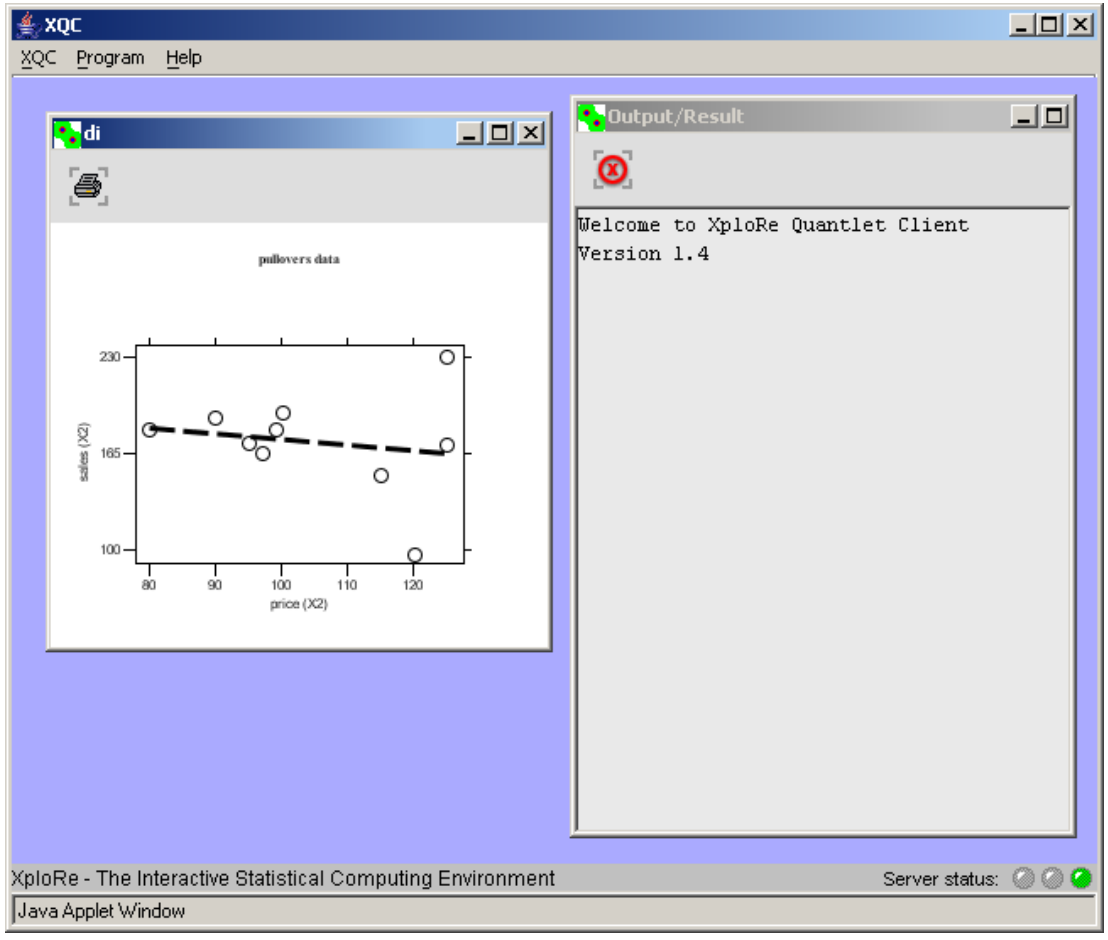
Applied Multivariate Analysis - Online



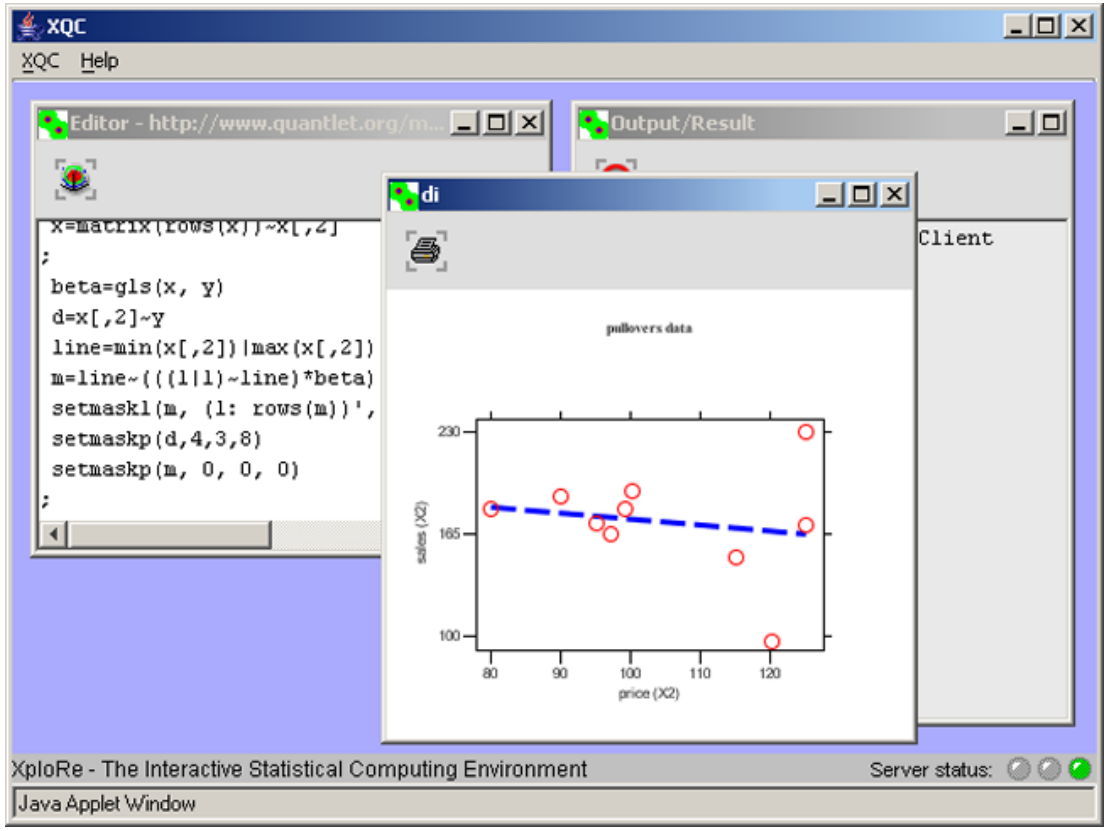
HTML-page for Electronic Examples



Interactive Example for MVA - Run



Interactive Example for MVA - Edit



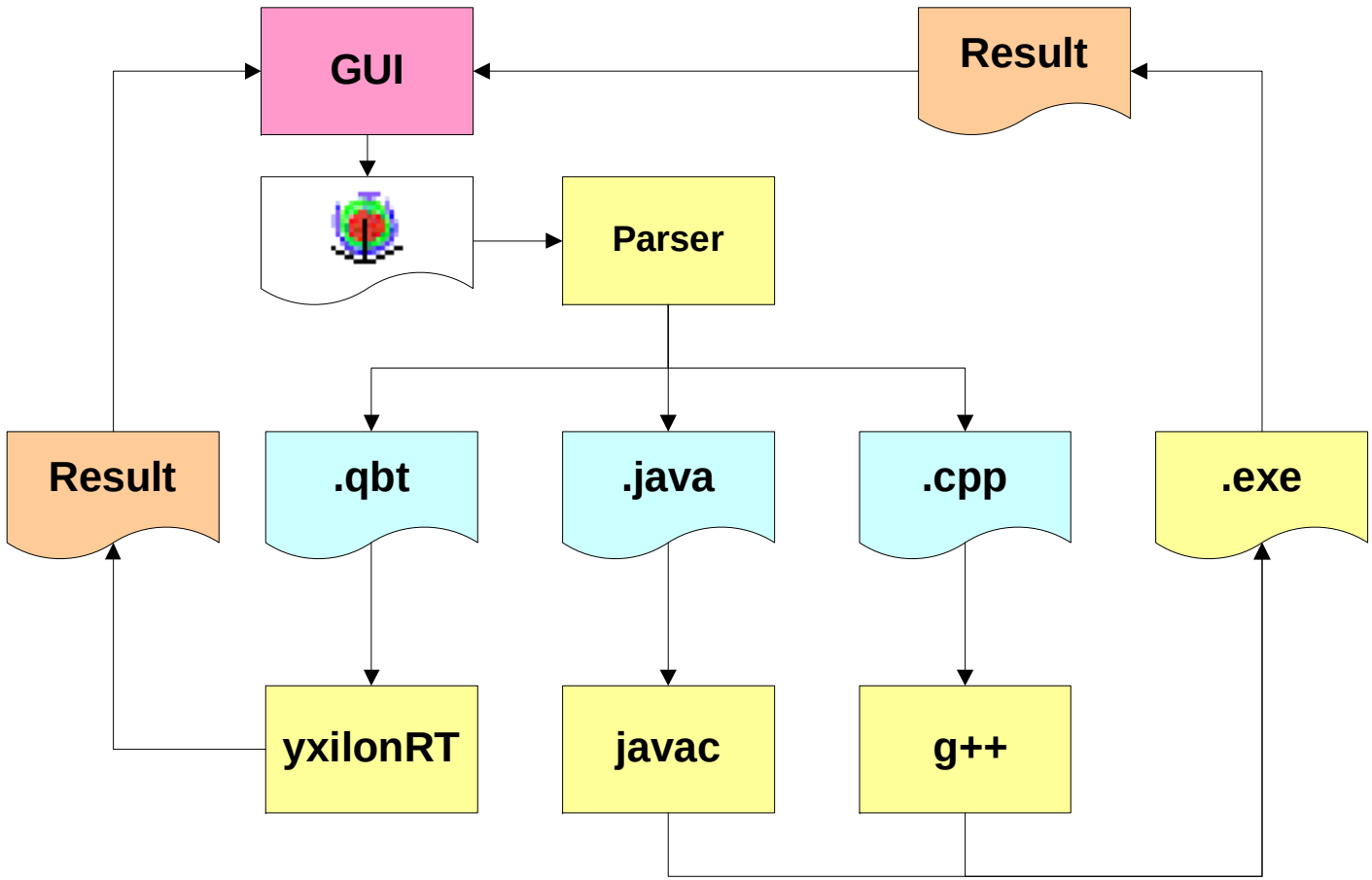
XploRe

- developed by Humboldt-Universität zu Berlin and MD*Tech
- C-style syntax, procedural approach
- available as batch, standalone and Client/Server on Win32 and Unix/Linux
- strong focus on non-parametric and quantitative finance methods

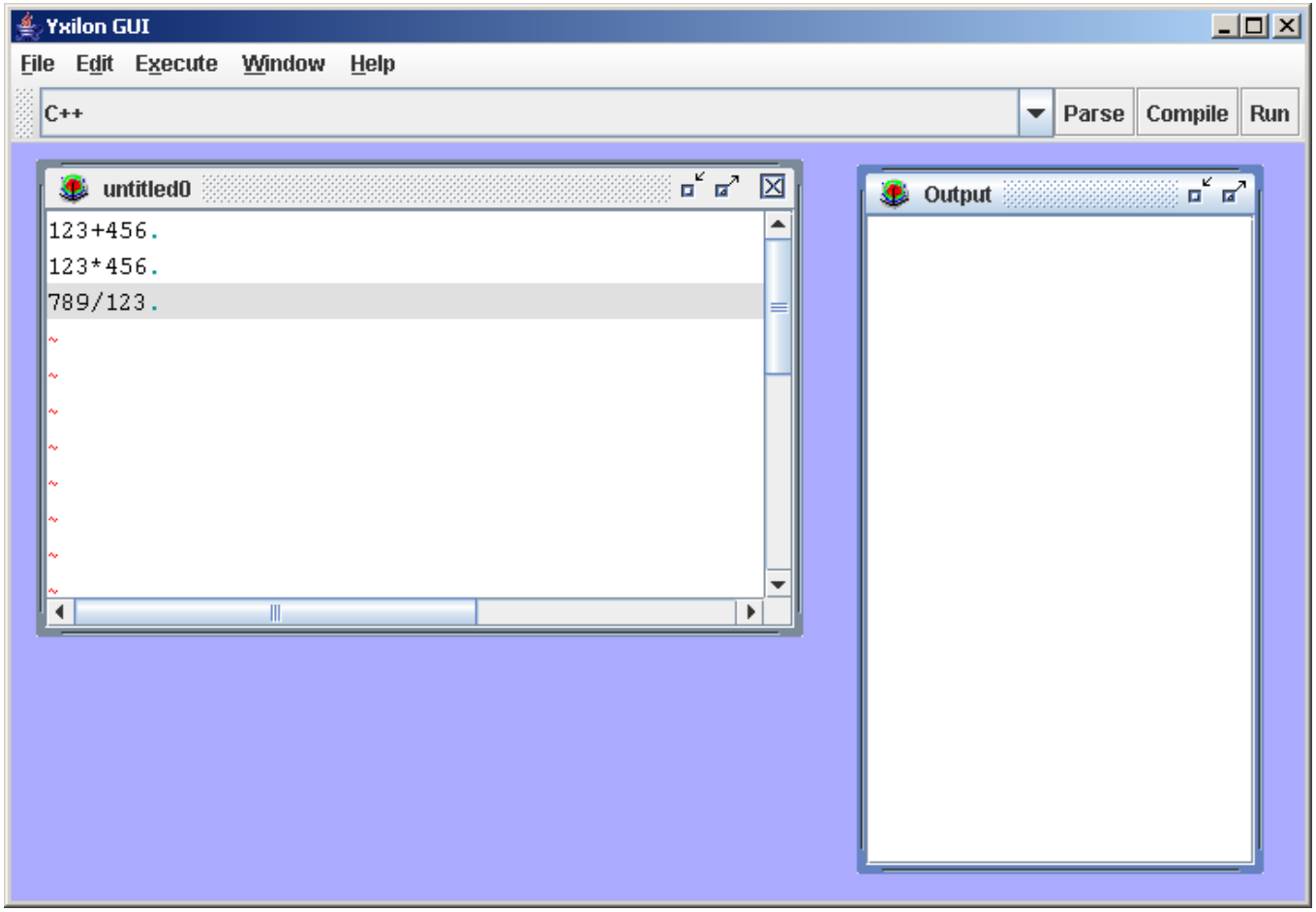
Yxilon

- Multiple front-ends: commandline, GUI, embedded into Excel
- Extensibility on language and native core level
- Read/write data and to calculate across networks
- Support for databases and interactive graphics
- Inclusion of existing code (`C`, `Fortran`, XploRe)

Compilation



Yxilon Screenshot



Technical Limitations

- *e*-tools need software architecture
  - Easy to handle and less powerful?
  - Powerful and complex
  - strategic decision, trade-off must be found
- MM*Stat: HTML, CSS and JavaScript are browser-dependent
- JAVA platform-independent, only intersection of functionality

Psychological Barriers

- computer knowledge among students still diverse
- psychological barriers to use *e*-lements for learning
- 90s: hype to teach online failed
- Do students want to learn online?

Educational Limitations

What cannot be taught via *e*-lements?

- complex data analysis has several steps
  - explorative and descriptive analysis
  - ANOVA or PCA
  - regression
  - statistical tests
- single steps are teachable well
- 'big picture' may get lost
- Statistical thinking cannot be taught

References


Borak, S., Härdle, W., Lehmann, H.(2005)
Working with the XQC
in *Statistical Tools for Finance and Insurance*
editors: Cizek, P., Härdle, W., Weron, R., Springer Verlag


Chambers, J. and Lang, D. T. (1999)
$\hat{\Omega}$– A Component-based Statistical Computing Environment
Proceedings of the 52nd Session of the ISI, Helsinki


Fujiwara, T., Ikunori K., Nakano, J., Yoshikazu, Y. (2000)
A Statistical Package Based on Pnuts
In: COMPSTAT. Proceedings in Computational Statistics, Physica Verlag


Guril, Y., Klinke, S., Ziegenhagen, U. (2005)
Yxilon – a Modular Open-Source Statistical Programming Language
In: Proceedings of the 55th ISI, Sydney


Mori, Y., Yamamoto, Y. and Yadohisa, H. (2003)
Data-oriented Learning System of Statistics based on Analysis Scenario/Story
Bulletin of the International Statistical Institute (ISI)


Müller, M., Rönz, B., Ziegenhagen, U. (2000)
The Multimedia Project MM*Stat for Teaching Statistics
In: COMPSTAT. Proceedings in Computational Statistics, Physica Verlag


Tukey, T. (1965)
The Technical Tools of Statistics
American Statistician 19, 23-28.