Teaching Statistics with XploRe*

Marlene Müller
Institute for Statistics and Econometrics, Humboldt University Berlin
Spandauer Str. 1, D–10178 Berlin, Germany

marlene@wiwi.hu-berlin.de, http://www.wiwi.hu-berlin.de/∼marlene

January 14, 1998

*To appear in: *Maths&Stats Newsletter*, CTI Statistics (http://www.stats.gla.ac.uk/cti/).
The research for this paper was supported by Sonderforschungsbereich 373 at the Humboldt-University Berlin. The paper is printed using funds made available by the Deutsche Forschungsgemeinschaft.

1 Overview

XploRe is an interactive computational environment for statistics. The aim of XploRe is to provide a full, high-level programming language with tools for user interaction and dynamic graphics. A central aspect is the possibility to equally use it within a local network or the internet. Therefore, XploRe comes in several flavours:

(1) Generic versions are available for Unix/X11 (Solaris/Sparc and Linux/PC) and under MS Windows (95/NT for PC).

(2) A Java client version is available, to be used with a XploRe server running under Unix. The server might run on a remote machine. The XploRe Java client runs under Java 1.1. Virtual Java machines are available for a wide number of operating systems.

(3) A Java applet version can be used from the XploRe home page http://www.xplore-stat.de which provides access to XploRe from any WWW browser that supports Java applets.

(4) A CGI interface version can be used from the XploRe home page http://www.xplore-stat.de access to XploRe from any WWW browser that supports forms.

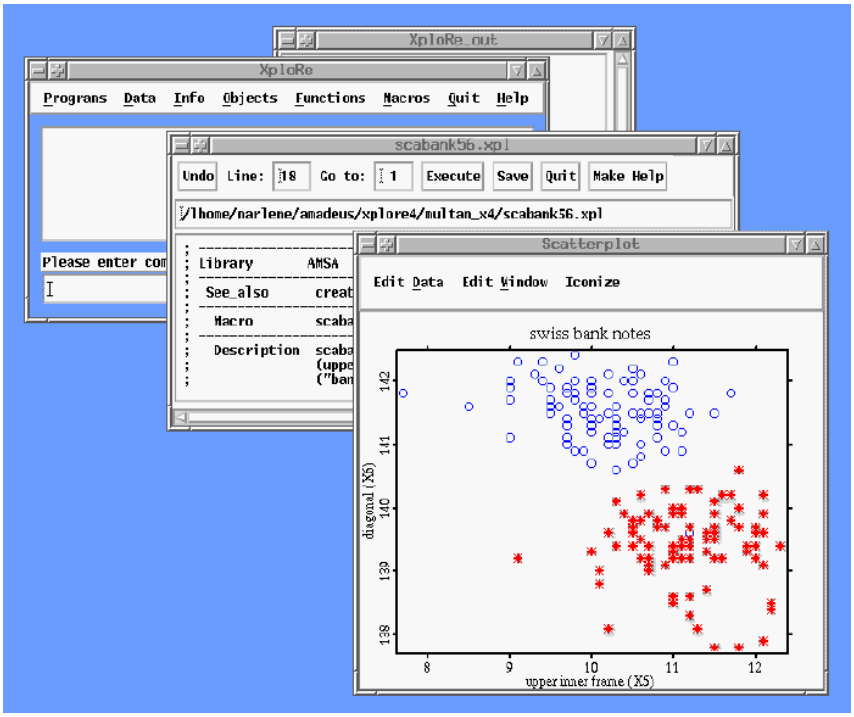

Figure 1: Screenshot of the generic XploRe version

To use the generic version (1) or the Java client version (2), a local copy of the software needs to be present on the users computer. To use the Java applet version (3) or the CGI version (4) only a WWW browser (providing java applets/forms) and an internet connection are necessary. This makes XploRe in particular interesting for teaching: Since standard WWW browsers offer both java applets and forms, students have an easy access

to XploRe from their home PC or from a PC pool at university or even an internet caf'e. Hence, students are not required to buy/download the software and can use it from almost everywhere. Additionally, the XploRe help system is provided in HTML and can be browsed with any common WWW browser.

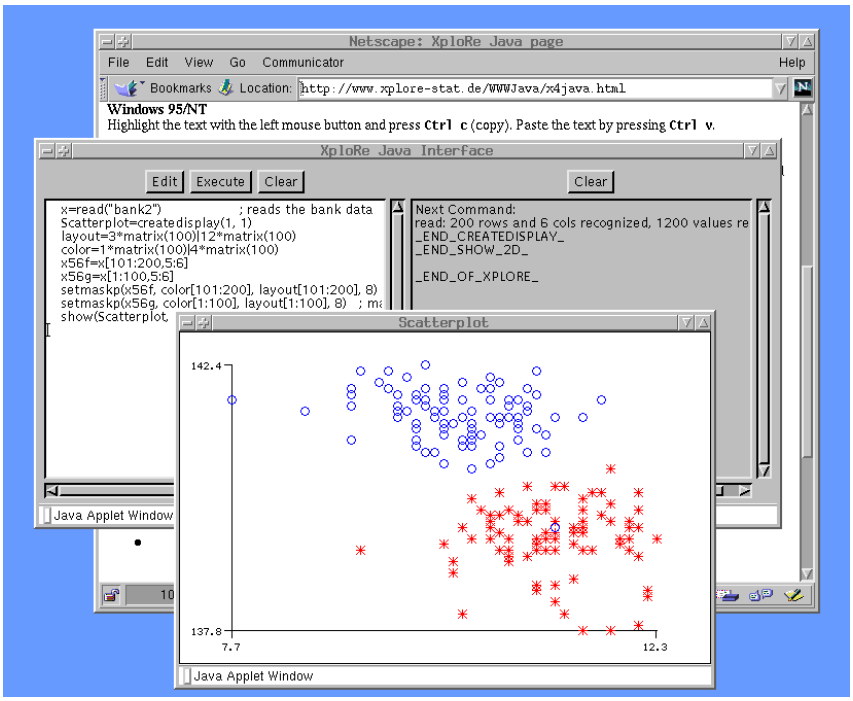

Figure 2: Screenshot of the XploRe Java version

XploRe provides a procedural language, which allows the user to write procedures or functions, like in Splus or Gauss. All features of an high-level language like recursion, local variables, loops, and conditional execution are available. It is an interpreted language which allows an immediate execution of XploRe commands or macros.

In the following, a number of examples will be presented that give an idea how XploRe is used to support and complement courses in introductory and advanced statistics. All these examples can be downloaded from the given WWW pages. Most of these examples can be directly run from the Java client or the Java interface, although it should be mentioned that not all interactive functionality is available Java version. Also, it should be mentioned, that using XploRe for teaching is still in a starting phase, not all the possibilities are uitilized yet.

The Institute for Statistics and Econometrics (http://wotan.wiwi.hu-berlin.de) provides a WWW page with teaching material in HTML, PostScript or PDF (http://wotan.wiwi.hu-berlin.de/statistik/lehrmaterial/statmat.html). These materials can be downloaded by students. Accompanying XploRe-macros can be either downloaded or directly executed via the XploRe Java interfaces.

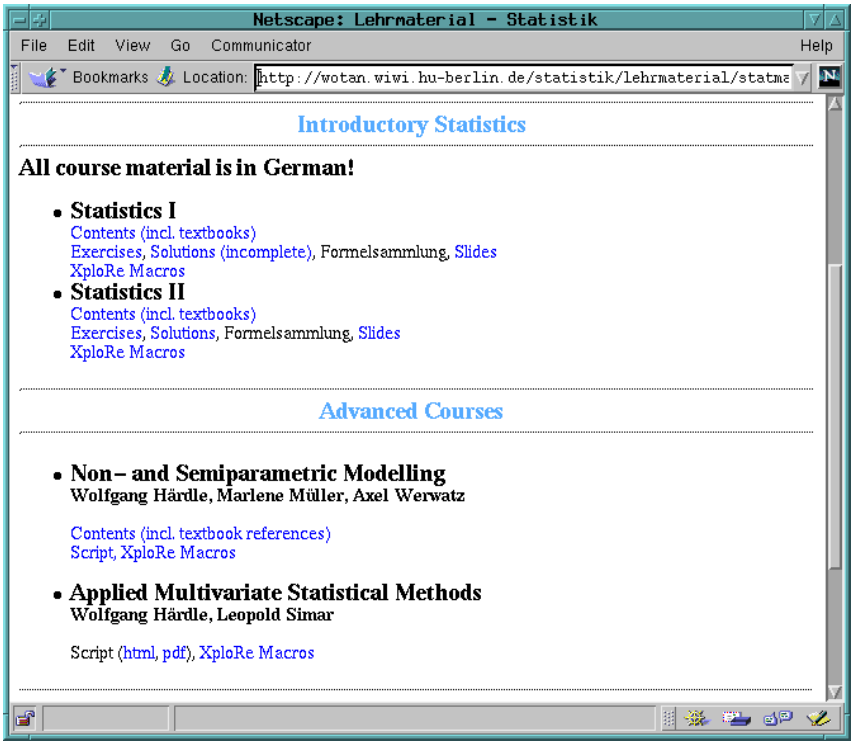

Figure 3: Course material available by WWW

2 Teaching introductory statistics

Due to the fact, that introductory statistics courses are intended for first and second year students, all course material is in German, whereas a part of the XploRe-macros is provided in both languages. Practically all material for the introductory courses is available via WWW (http://wotan.wiwi.hu-berlin.de/statistik/lehrmaterial/statmat.html).

In introductory statistics, only a few computer-based examples are used in the course. Computer-assisted teaching is meant to complement the course and not (yet) an integral part of the course. This is reflected by the facts, that (a) only a few XploRe routines are provided and (b) these routines are primarily presented by the teacher and not necessarily used by the students themselves.

The first part of the introductory statistics course (Statistics I) mainly deals with descriptive statistics. This covers in particular: different types of statistical variables, their graphical representation and characteristics (mean, variance, median, quantiles, correlation etc.).

Example 2.1: Descriptive Statistics (Credit.xpl)

The first XploRe example gives an interface for a standard set of these methods. The XploRe macros analyses a credit data set of 25 individuals (subsample from Fahrmeir, Tutz, 1994). The dataset has five variables: personal id, credit worthiness, purpose of credit, monthly payments (from 1=low to 4=high) and amount of credit. The variables represent different types: binary, discrete, ordinal and continuous.

The purpose of the macro is to show the different techniques for explorative analysis.

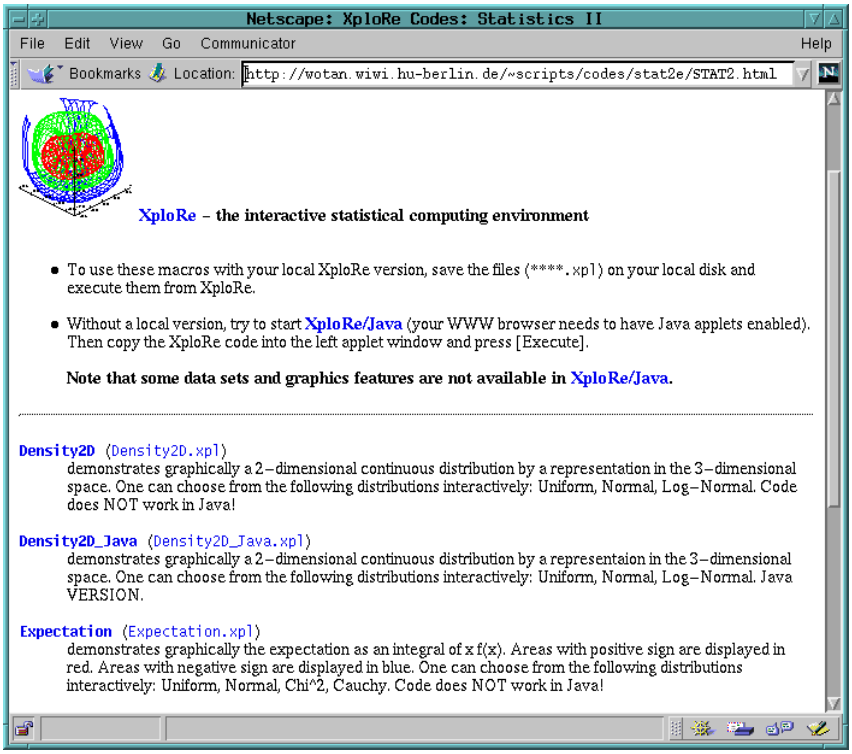

Figure 4: XploRe macro list for Statistics II

Different tools can be used, for example a barchart can be computed for each variable, although it may not be reasonable in all cases. Figure 5 shows an explorative analysis of the continuous variable amount.

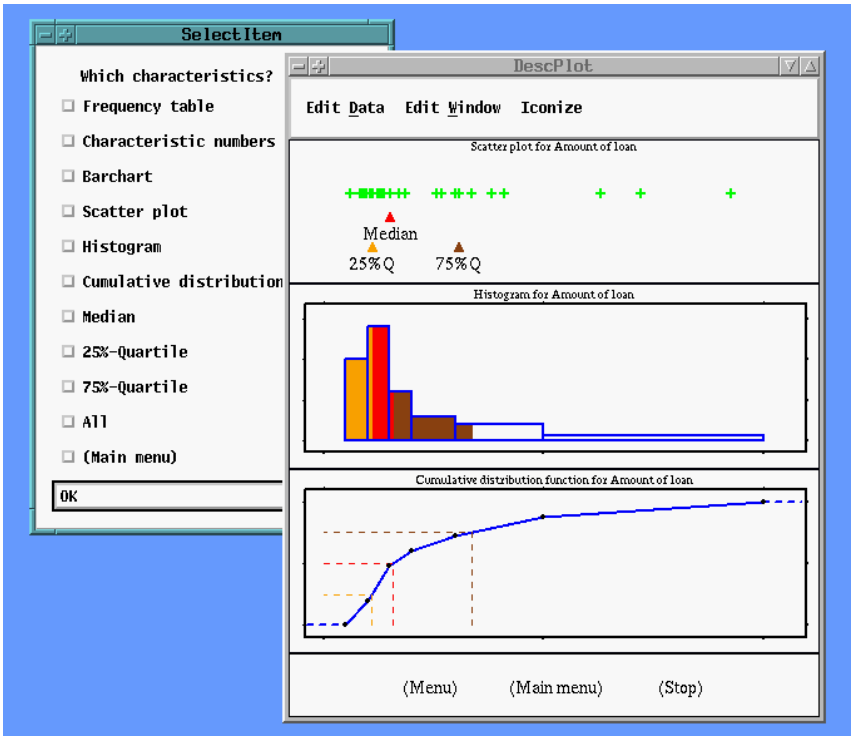

Figure 5: Descriptive Statistics

The second part of the introductory statistics course (Statistics II) is devoted to probability calculus and theoretical statistical concepts. Main issues are random variables, sampling theory, parameter estimation and hypothesis testing.

Example 2.2: Expectation (Expectation.xpl)

The XploRe example in Figure 6 demonstrates graphically the expectation as an integral of x f(x). Areas with positive sign are displayed in red and areas with negative sign are displayed in blue. One can choose following distributions from a menu: Uniform, Normal, Chi-square, Cauchy. For all these distributions also the parameters can be modified, such that it is possible to display the effect of moving the distribution along the horizontal axis or to exploit the effect of changing the scale parameter.

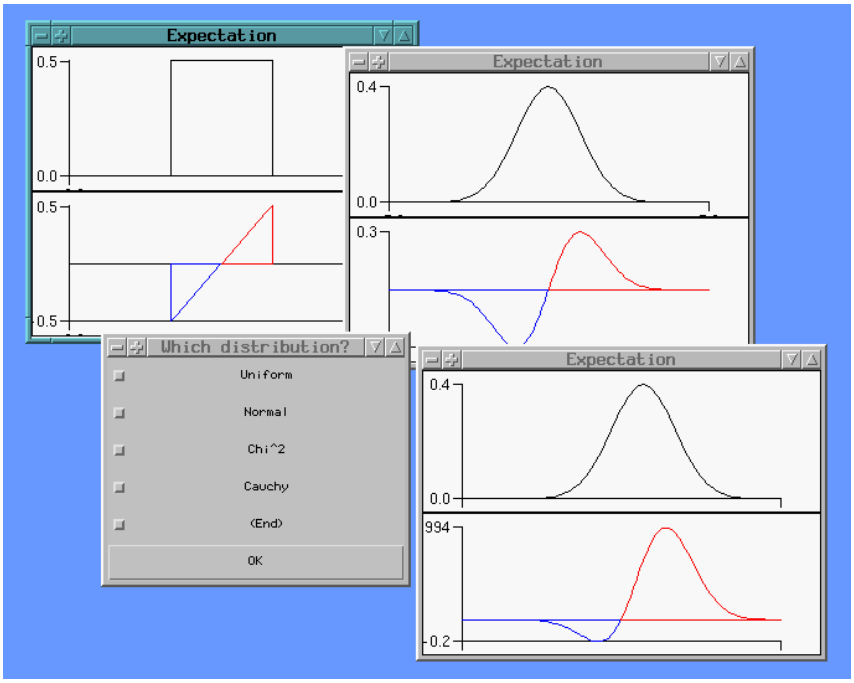

Figure 6: Graphical demonstration of expectation

3 Advanced statistics courses

Advanced courses, in which XploRe is used by students, cover multivariate statistical methods, non- and semiparametric modelling, option pricing and interactive statistics. A collection of XploRe macros together with a course script in electronic form is available for the courses "Applied Multivariate statistical Analysis" and "Non- and Semiparametric Modelling".

The macro collection for both courses consists of routines that can be used independently from each other. When using the electronic version of the course script (HTML version) it is possible to directly access the XploRe routines used for the examples, since they are referred by an link.

In contrast to the introductory courses, the students themselves use the XploRe for laboratory exercises as well as for their homework. As before, everything is provided on

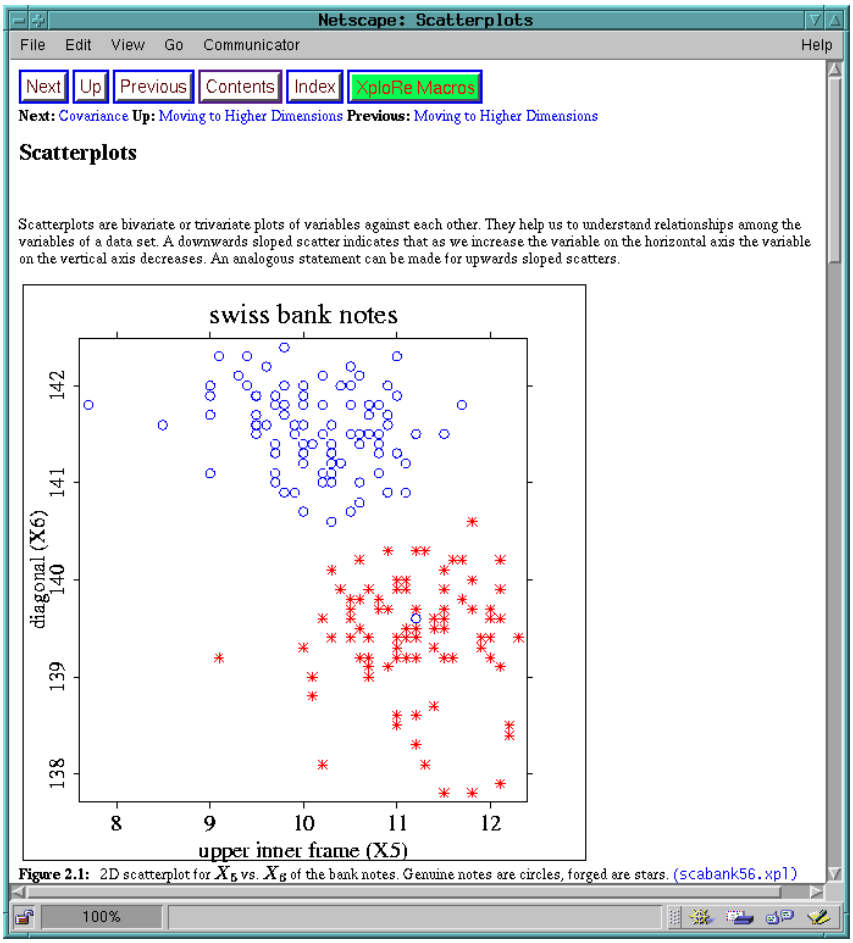

Figure 7: Script and link to XploRe macro

the WWW server (http://wotan.wiwi.hu-berlin.de/statistik/lehrmaterial/statmat.html) and macros can be used and tried out independently from each other. The XploRe codes contain comments which make it easy to change parameters or input data sets. A short example is the macro scabank45.xpl which shows a scatterplot for two variables.

```
; ------------------------------------------------------------------
; Library       AMSA
; ------------------------------------------------------------------
;  See_also      read createdisplay show
; ------------------------------------------------------------------
;   Macro        scabank45
; ------------------------------------------------------------------
;   Description  computes a two dimensional scatterplot of
;                X4 vs. X5 (upper inner frame vs. lower) of
;                the Swiss bank notes data ("bank2.dat")
; ------------------------------------------------------------------
   x=read("bank2")                    ; reads the bank data
   x1=x[,4:5]
   Scatterplot=createdisplay(1, 1)    ; creates display
   show(Scatterplot, 1, 1, x1)        ; 2D plot of variables 4 and 5
```

Three examples from multivariate analysis and nonparametric function estimation will make this more clear. All these examples can be used from the XploRe Java client or the Java applet which is available with an internet connection.

Example 3.1: Multivariate Statistics of Swiss Bank notes
(andcur.xpl, boxplot6.xpl, scabank56.xpl, pcabank.xpl)

A running example in the multivariate analysis course
(http://wotan.wiwi.hu-berlin.de/~scripts/codes/sma/AMSA.html) is the Swiss bank note data set (Flury, Riedwyl, 1988) which consists of 200 bank notes. The data contain two subgroups: 100 bank notes are genuine and 100 are forged. The problem connected with these data is to find a discrimination rule to separate the two groups. Throughout the course the data are used to explain the application of graphical and multivariate techniques.

The first part of the multivariate analysis course is devoted to the graphical exploration of the bank notes data. Figure 8 shows the result of 4 XploRe macros which produce a boxplot of variable X6 ("diagonal of bank note"), a scatterplot of variables X5 ("lower frame size") and X6, Andrews curves of 5 genuine and 5 forged bank notes and a principal components analysis of the whole data. For easier interpretation the data are coloured differently in both subgroups.

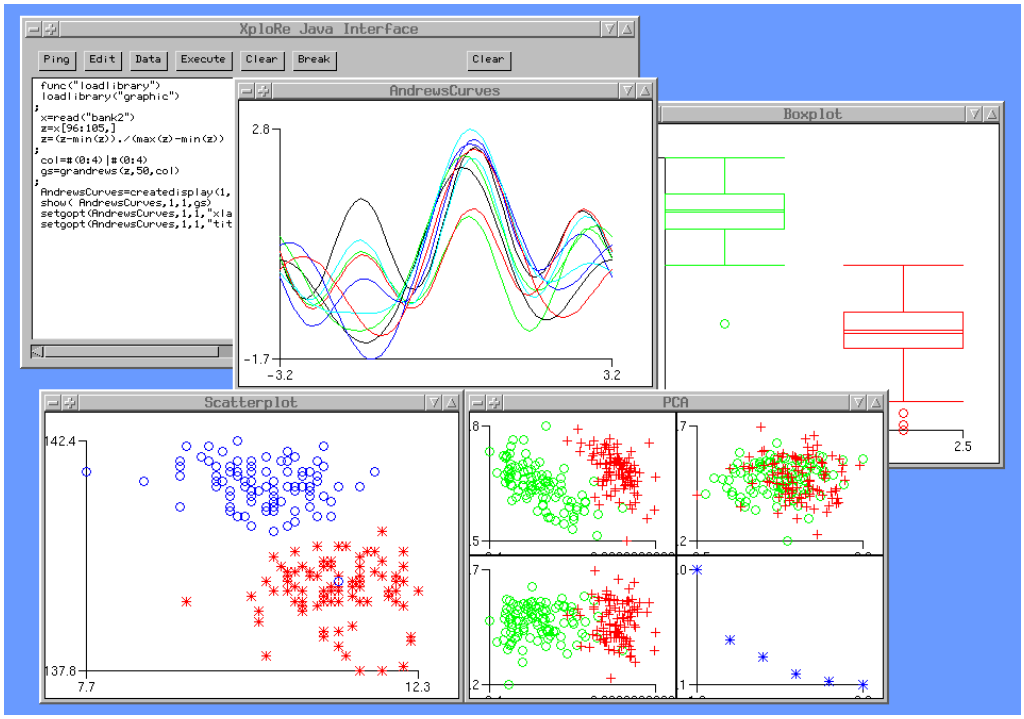

Figure 8: Graphical exploration of Swiss bank notes data

Nonparametric density and regression estimation, in particular by kernel methods, is a part of the course on non- and semiparametric models. The concept of this macro collection (http://wotan.wiwi.hu-berlin.de/~scripts/codes/spm/NSPM.html) is similar to this for multivariate analysis.

Example 3.2: Interactive Averaged Shifted Histograms (InteractiveWARPing.xpl)

An easy to explain method for nonparametric density estimation is given by the averaged

shifted histogram (Scott, 1992). Figure 9 shows the interactive application of the averaged shifted histogram (ASH) on netincome data (subsample for UK family expenditure survey 1973). The parameters of the ASH, the binwidth h and the number k of histograms to be averaged, can be chosen from a menu. The result is a double display, such that always two ASHs with different parameter settings can be compared.

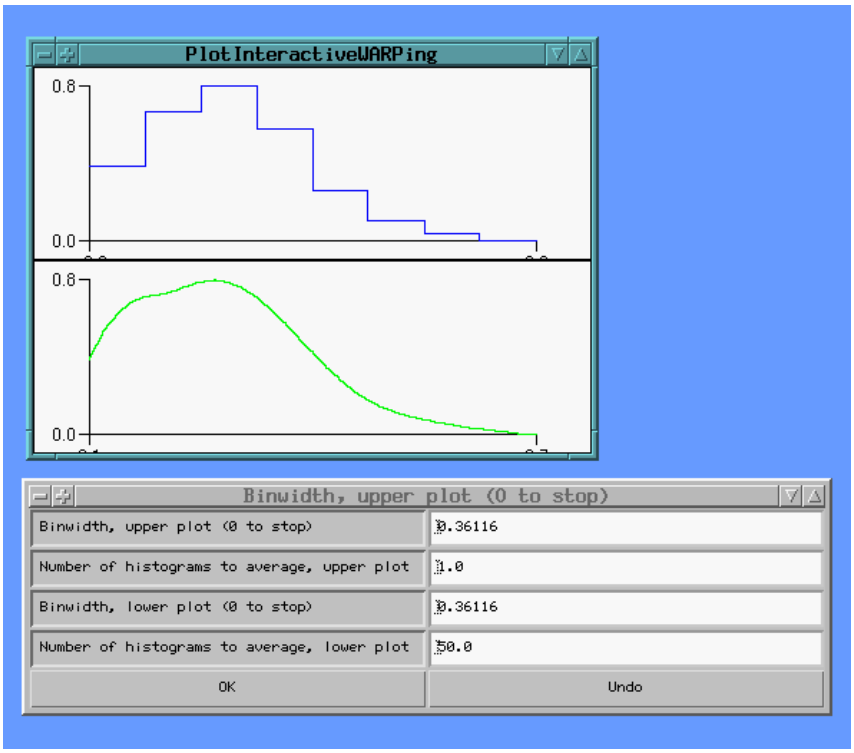

Figure 9: Comparing averaged shifted histograms w.r.t. their parameters

Example 3.3: Interactive Regression Estimation (Interactive Regression.xpl)

An important topic in nonparametric statistics is kernel regression estimation. The process of bandwidth choice is suitable for an interactive routine as well. Figure 10 displays the result of an user chosen regression estimate for the regression of food share on household netincome (subsample for UK family expenditure survey 1973). During the process of bandwidth choice, the cross-validation curve (blue) is displayed on the lower panel, with blue bullets indicating all chosen bandwidths and a red bullet indicating the currently used bandwidth.

References

L. Fahrmeir, G. Tutz (1994): Multivariate Statistical Modelling Based on Generalized Linear Models, Springer.

B. Flury, H. Riedwyl (1988): Multivariate Statistics. A practical Approach, Chapman and Hall, London.

W. Härdle, L. Simar (1998): Applied Multivariate Statistical Methods, Course Script, Institute for Statistics and Econometrics, Humboldt-University Berlin.

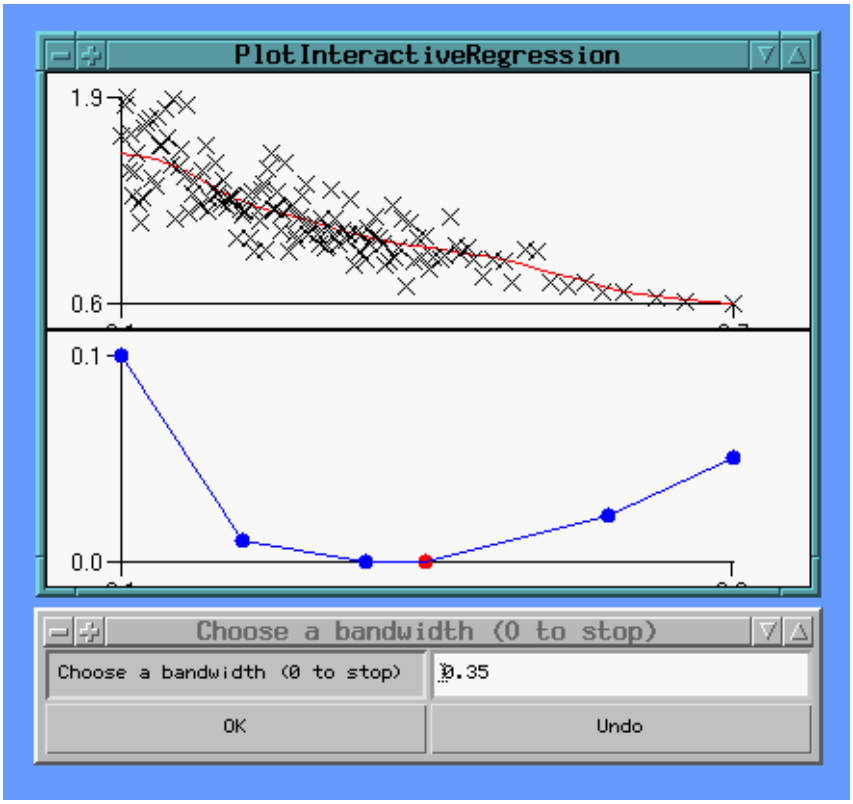

Figure 10: Fitting a kernel regression interactively

W. Härdle, M. Müller, A. Werwatz (1998): Non- and Semiparametric Modelling, Course Script, Institute for Statistics and Econometrics, Humboldt-University Berlin.

S. Schmelzer, T. Kötter, S. Klinke, W. Härdle (1996): A New Generation of a Statistical Computing Environment on the Net, COMPSTAT 96, Barcelona

D. Scott (1992): Multivariate Density Estimation: Theory, Practice, and Visualization, John Wiley and Sons.

G. Teyssière (1998): XploRe 4.0, A nonparametric statistical Package, review, to appear in Journal of Econometrics

WWW-Links and further Information

XploRe
http://www.xplore-stat.de

Course Material and XploRe Macros
http://wotan.wiwi.hu-berlin.de/statistik/lehrmaterial/statmat.html