Teaching Statistics With R: Beyond E-Learning Systems

Matthias Templ, Gerlinde Dinges

Department of Statistics and Probability Theory
Vienna University of Technology
& Statistics Austria

Boku, January. 16, 2009

Content

Teaching Statistics in Statistics Austria

- **Methodology 1** (4×3 hours): Motivation to Statistics, Basics, Descriptive Statistics, Probability Theory, Indices, Data Visualization.
- **Methodology 2** (5×3 hours): Complex Sampling Theory, Calibration, Point and Interval Estimation, Tests, Quality Aspects, Covariance Estimation.
- **Methodology 3** (5×3 hours): Regression, Time Series Analysis, Saisonal Adjustment, Imputation, Editing, Statistical Matching.
- **Statistical Disclosure Control** (1×4 hours): Remote Execution and Remote Access, Microdata Protection, Protection of statistical tables.

Facts

- course participants with (very) different levels of education in statistics.
- course participants with (very) different levels of motivation.
- various topics to teach but within a (very) limited time.

$\rightarrow$ Teaching Statistics **independently** from the statistical background of the participants.

Objectives

- Each course participant **is going to** contribute **interactively** to the course.
- Making theoretical concepts understandable. Interactive **visualisation** of any mathematical expression.
- Embedding of information about our participants into the course material to **attract interest**.
- Teaching **Robust Statistics** but also explain standard concepts (but **not** vice versa).

Virtual Book

Use of Data from our Participants

- Participants should identify their **own** data in examples to attract interest. (Who is the outlier?)
- **Collection** of data from the participants via E-questionnaire within the course.
- Embedding these data **automatically** in interactive examples and in the course slides in **real time**.

Visualization of Anything

- For supporting the **explanation** of theoretical concepts.
- The participants should not only **hear** about statistics, they should **see** it in order to **understand** it.
- We derive visually **step-by-step** fundamental concepts but also graphics like QQ-plots, etc. (all in all ≥ 500 plots)

Robustification of Nearly Anything

Teaching Robust Statistics

- as natural as possible because
    - virtually all of the data in Statistics Austria include **outliers**.
    - even course participants with good knowledge in Statistics don't know anything about it. (again: attract interest)
- to understand problems arising when standard techniques are applied.

Active Contribution of our Participants

- Don't waste time with calculations as exercises.
- Don't use one particular statistical software for exercises. (why should an SAS user use R and vice versa?)
- We considered the use of e-learning systems available through the web but we conclude that none fits within our ideas.

The Interactive Knowledge-Check (TGUI)

Former: No active contribution of the course participants. (any questions? → silence)

Development of a specific interactive Knowledge-Check

The **TGUI**:

- **About**: Recapitulation of the previous topic by multiple choice questionnaires or interactive visualizations.
- **Why**: Previous topics should be strengthend. Next topic should be motivated. Attract interest.
- **How**: Interactive (and anonymously) after every 20 minutes (presentation) via mouse clicks and sliders.
- **Advantage**: Feedback for both, the course presenters and the participants. How many participants have successfully solved the examples? Should we repeat something? What?
  No time-consuming examples which must be solved manually.
- Interactive gaming without being stressed by exams - **attract interest**.

Software

- LATEX is a must, not only because of better fonts for mathematical formulas.

- R is a must, not only because of its power to generate high-quality plots.

- Sweave to combine both, R and LATEX for the dynamical generation of reports in real time.

- JGR (or GGOBI) for various interactive plots.

- ImageMagick to produce animimations.

- tcltk (or RGTK2) to implement a graphical user interface (GUI).

Input-Output
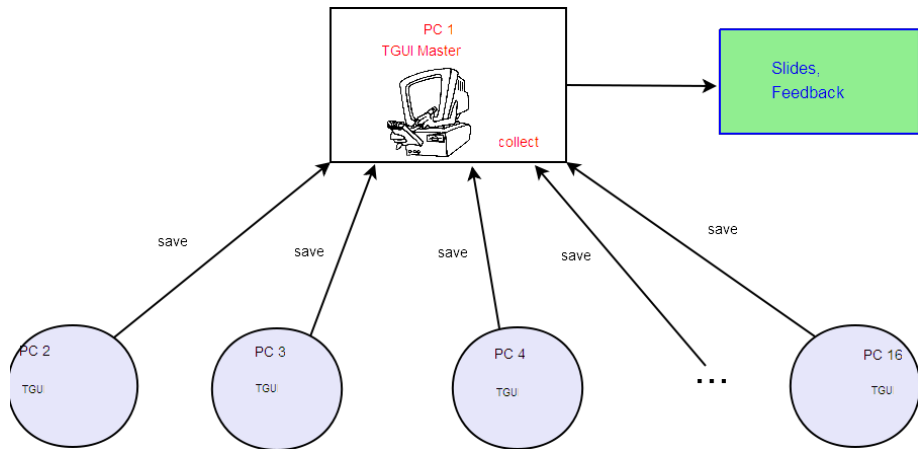
After the E-Questionnaire

[1.1] Am Weg zur linearen Regression

Ihre Datenmatrix:

```
Nr Pseudonym Geschl Größe Gew Alter Quelle
1     Obelix  Männl   187 125    51   Kurs
2      Karli  Männl   180  75    28   Kurs
3       Susi  Weibl   170  65    23   Kurs
4         yt  Männl   183  76    28   Kurs
5       Lisa  Weibl   168  54    32   Kurs
6      Frosch Weibl   170  61    40   Kurs
7    Gilbert  Männl   164  62    34   Kurs
8      Maria  Weibl   154  59    54   Kurs
9       Anid  Männl   183  71    27   Kurs
10   nobody1  Männl   175  90    42   Kurs
11     Kurti  Männl   172  59    27   Kurs
12   Karfiol  Weibl   160  55    36   Kurs
13      maxi  Weibl   168  50    35   Kurs
14    Moritz  Männl   180  92    33   Kurs
15     Quasi  Männl   173  68    28   Kurs
16     Linus  Männl   183  79    46   Kurs
17  Garfield  Weibl   167  57    41   Dummy
18 IchAnonym  Weibl   178  68    39   Kurs
```

Quelle: ST.AT, 15. Mai 2005

Matrix-Schreibweise:
n Beobachtungen (Zeilen) und p Variablen (Spalten)

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \ldots & d_{1p} \\ d_{21} & d_{22} & \ldots & d_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ d_{n1} & d_{n2} & \ldots & d_{np} \end{pmatrix}.$$

Notation:
d_{11} ... Wert in Zeile 1, Spalte 1
d_{35} ... Wert in Zeile 3, Spalte 5
$d_{.1}$ ... 1. Variable (1. Spalte)
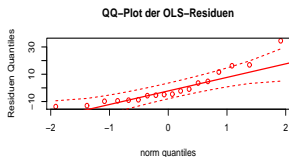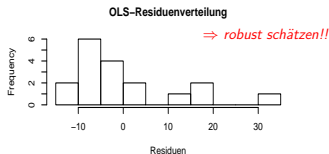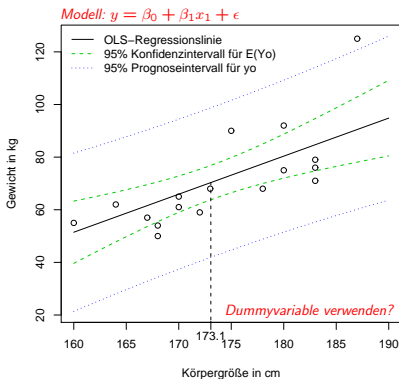$d_{3.}$ ... 3. Beobachtung (3. Zeile)

[2.1] OLS: Diagnostik

Beispiel 5/1: Graphische Verfahren zur Analyse der Residuen e_i

Die *Residuenverteilung* vermittelt einen ersten Eindruck über die Güte der Prognosefunktion

Quelle: ST.AT, 15. Mai 2005

Examples from the TGUI

Online . . .

- ST03 - 13.1
- ST03 - 14.1
- ST03 - 19.1
- ST04 - E1
- ST03 - 8b
- ST05 - OLS
- ST05 - gp

Examples from the TGUI

- Education in Statistics is essential for employees of Statistics Austria.
- The education must be practical oriented and must **entertain** the employees. Note that without any possibility of giving exams for (often prior unmotivated) participants one must give a strong motivation.
- The **feedback system** is essential to check if the topics have been understood by the participants.
- All in all, our concept consists of a whole **system**, consisting not only of the TGUI.
- This concept can be applied anytime to present any topic, but a computer lab is necessary. We assume that in future for every student a computer is available during the lecture, so the concept is going to be **general**.
- The concept works - the evaluation of the course is almost always.