

Motivation zur Statistik – Computergestützt lernen in der Statistik Austria

Gerlinde Dinges¹ und Matthias Templ^{1,2}

¹FB Methodik, Statistik Austria, Wien

²Statistik und Wahrscheinlichkeitstheorie, Technische Universität Wien

(Die Autoren haben zu gleichen Teilen an dieser Arbeit beigetragen)

Abstract: To satisfy the ongoing scientific progress Statistics Austria provides the opportunity to acquire, brush up or deepen statistical knowledge in training courses on current basic statistical methodology. In order to convey the mathematical contents of the courses and the complex interrelationships in the most comprehensive way, the authors of this article developed a special statistical training programme. The programme is computer based and has been successfully applied as part of the in-house employee training since 2005. The learning system presented here is designed for modern, computer-assisted teaching and provides an insight to other lecturers or institutions to get further ideas or alternative concepts on how to structure basic teaching courses.

Zusammenfassung: Dem ständigen wissenschaftlichen Fortschritt Rechnung tragend bietet Statistik Austria Interessierten die Möglichkeit, in methodischen Grundlagenkursen aktuelles statistisches Basiswissen zu erlangen, oder bereits erworbene Kenntnisse aufzufrischen und zu erweitern. Um mathematische Kursinhalte und komplexe Zusammenhänge möglichst kompakt und für alle verständlich zu transportieren, haben die Autoren des Artikels ein spezielles computergestütztes Statistiktraining entwickelt, welches seit 2005 im Rahmen hausinterner Mitarbeiterschulungen erfolgreich Anwendung findet. Das Lernsystem ist für die moderne computergestützte Lehre konzipiert und soll durch den hier präsentierten Einblick auch anderen Lehrenden oder Institutionen als alternative Idee oder mögliches Konzept einer effektiven Kursgestaltung vorgestellt werden.

Keywords: Computergestütztes Lernsystem, Datenvisualisierung.

1 Einleitung

Als moderne Informationsgesellschaft leben wir im Zeitalter der Statistik. Tagtäglich werden wir über die Medien mit Informationen unterschiedlichster Qualität konfrontiert, zu deren Bewertung und Verständnis statistische Grundkenntnisse von Vorteil sind¹. Auch im beruflichen Alltag gewinnt statistisches Wissen mehr und mehr an Bedeutung – nicht zuletzt durch immer leistungsfähigere Computer- und Softwaresysteme. Weniger erwünscht dabei ist, dass durch einfach bedienbare Softwareprodukte auch manch methodisch Unkundiger zur ‘Datenanalyse’ verleitet wird. Per Knopfdruck spucken Programme eindrucksvolle Zahlen aus, die ohne entsprechende Statistikkenntnisse oft falsch interpretiert

¹Vgl. <http://www.stat4u.at>, ein Projekt des Instituts für Scientific Computing der Universität Wien.

werden oder von vornherein bereits völlig unsinnig sind. Vorrangiges Ziel einer statistischen Grundausbildung sollte deshalb auch sein, zuerst jene solide und wichtige Basis zu liefern, die für den korrekten Einsatz spezieller Statistik- und Analyseprogramme Grundvoraussetzung ist. Die Handhabung der Software selbst ist danach rasch erlernt – sofern sich der Anwender auf das einfache Bedienen einer graphischen Benutzeroberfläche beschränkt, was bei fehlenden Programmierkenntnissen in der Regel der Fall ist.

Zur laufenden internen Mitarbeiterschulung bietet Statistik Austria (STAT), neben fachstatistischen Kursen, auch statistisch/methodische Grundlagenkurse an, für deren Besuch kein spezielles Fachwissen erforderlich ist. Den STAT-Methodikern stellte sich damit die Aufgabe, die 2005 erstmals angebotenen Grundlagenkurse so zu gestalten, dass mathematische Kursinhalte und komplexe Zusammenhänge möglichst kompakt, und dennoch für alle verständlich transportiert werden. Moderne Methoden wurden eingesetzt und ein spezielles computergestütztes Statistiktraining entwickelt, welches unabhängig von den individuellen Softwarekenntnissen der Kursbesucher eingesetzt werden kann. Über Umsetzung, Methodik und Erfahrungen wird im vorliegenden Beitrag berichtet.

2 Aufgabenstellung

Die von den STAT-Methodikern² zur Mitarbeiterschulung angebotenen Grundlagenkurse sollten aufbauend gestaltet sein und dazu dienen, statistisches Basiswissen zu erlangen, oder dazu beitragen, bereits vorhandenes Wissen aufzufrischen und durch den Ausblick auf neue Techniken und Verfahren zu erweitern.

Die **methodischen Grundlagenkurse** gliedern sich in

- *Methodische Grundlagen I* (4×3 Stunden): Motivation zu Statistik, Statistische Grundbegriffe, Deskriptive (robuste) Statistik, Wahrscheinlichkeitstheorie, Index- und Verhältniszahlen (Exkurs), Datenvisualisierung.
- *Methodische Grundlagen II* (5×3 Stunden): Stichproben-, Hochrechnungs-, Schätz- und Testverfahren, Qualität stat. Produkte, Zusammenhang zwischen Merkmalen.
- *Methodische Grundlagen III* (5×3 Stunden): lineare (robuste) Regression, Zeitreihen- und Saisonbereinigungsverfahren, Imputation und Datenbereinigung, Methoden der Datenverknüpfung, Statistische Geheimhaltung.

Der Kurs *Methodische Grundlagen I* sollte von jedem Interessierten, unabhängig von Ausbildung oder fachlicher Verwendung, besucht werden können. Es war deshalb auch zu erwarten, dass eine Kursgruppe sich eher inhomogen zusammensetzen wird. Die Palette kann vom statistischen Neuling bis zum ausgebildeten Mathematiker reichen – mit Wissen aus Studium oder Praxis. Zudem konnte von unterschiedlicher Motivation für den Kursbesuch ausgegangen werden, da es bei einem um Mitarbeiterschulung bemühten Unternehmen mitunter auch jemanden nicht nur durch eigenen Wissensdrang in einen Kurs verschlägt. Es galt also unterschiedliches Vorwissen und Motivation unter einen Hut zu bringen, wobei sich das unterschiedliche Vorwissen der Kursteilnehmer und die aus zeitlichen Gründen in sehr kompakter Form zu vermittelnden (umfangreichen) Inhalte für die Gestaltung der Grundlagenkurse zunächst als am schwierigsten erwiesen.

²Alois Haslinger, Thomas Burg, Markus Fröhlich, Gerlinde Dinges, Matthias Templ, Bernhard Meindl

Wie kann es nun gelingen unter diesen Rahmenbedingungen eine effektive Grundausbildung anzubieten?

3 Die Idee

Zentrale Idee war es, alle Kursteilnehmer von Beginn an (inter-)aktiv in den Kursablauf einzubinden. Um das Interesse aller Beteiligten zu wecken und das Verständnis der Materie zu fördern, sollten überdies theoretische bzw. mathematische Inhalte möglichst verständlich aufbereitet und weitestgehend visualisiert werden. Indem von den Kursteilnehmern selbst erstellte Daten direkt in die verwendeten/begleitenden Vortragsunterlagen eingebettet sind, sollte die allgemeine Aufmerksamkeit während eines Vortrages zusätzlich erhöht werden. Bei den Kursinhalten sollten zudem neben der klassischen Statistik auch (moderne) robuste Verfahren Berücksichtigung finden.

3.1 Virtuelles Statistikbuch

Zuerst waren gut aufbereitete Kursunterlagen zu erstellen. Die angebotenen Methodikkurse finden, seit 2005, 1-2 mal jährlich in entsprechend ausgestatteten Computer-Labs der Statistik Austria statt, wobei die Teilnehmeranzahl auf 16 Personen begrenzt ist. Jedem Kursbesucher steht während des Kurses ein eigener PC zur Verfügung. Während der Vortragende ein bestimmtes Thema erklärt, sind alle PCs auf den Hauptcomputer geschaltet. Beim Zuhörer scheint dadurch am Bildschirm vor sich ein *virtuelles Statistikbuch* auf, das Schritt für Schritt den Vortrag begleitet. Eigenes Arbeiten am PC ist während ein Thema erklärt und im Statistikbuch ‘geblättert’ wird nicht möglich. Das hat den Vorteil, dass sich alle Beteiligten in dieser Zeit voll auf das behandelte Thema konzentrieren können und auch kein Schäfchen versehentlich im *World Wide Web* ‘verloren geht’.

3.2 Selbstmotivation durch eigene Daten

Als sehr wichtig erachtet wurde es, dass sich die Kursteilnehmer in den demonstrierten Beispielen wiederfinden. Zur Erklärung von statistischen Grundbegriffen oder Merkmalszusammenhängen genügt bereits ein sehr einfacher Datensatz. Dabei sollten allerdings nicht irgendwelche, sondern von den Beteiligten selbst erstellte Daten verwendet werden. Die Datenerstellung erfolgt mittels kurzer elektronischer Befragung. Danach kann die eigene Dateneingabe beispielsweise als anonymer Punkt in einer zweidimensionalen graphischen Darstellung von Gewicht und Körpergröße am Bildschirm betrachtet werden. Sollte es der als Versuchsobjekt gewählte ‘Göttergatte’ sein, der den OLS-Schätzer nach oben verzerrt? Und, was wäre die Lösung des Problems – ‘Diät’, oder vielleicht doch *Robuste Schätzung*? Oft wird bei persönlichen Daten eifrig diskutiert, was nicht verwunderlich ist, denn jeder weiß ‘*die Ausreißer sind unter uns*’. Wer selbst keine Daten angeben möchte, für den wird aus einem bestehenden Pool automatisch ein Dummy gezogen. Motto soll schließlich sein, das Wesentliche auf möglichst kurzweilige und ungezwungene Art und Weise zu vermitteln. Natürlich werden in weiterer Folge auch komplexere Datensätze ganz ‘seriös’ analysiert. Und obwohl die explorative Herangehensweise an Daten im Vordergrund steht, werden auch einige mathematische Herleitungen besprochen.

3.3 Visualisierung von fast ‘Allem’

Die Visualisierung von fast ‘Allem’ spielt in den Grundlagenkursen eine wesentliche Rolle und soll dazu beitragen theoretische Inhalte besser zu vermitteln. Außerdem sind schön dargestellte mathematische Formeln (auf die natürlich nie vergessen wird) oder gut aufbereitete Grafiken übersichtlicher und somit an sich bereits ‘leichter verdaulich’. Die Teilnehmer der Grundlagenkurse sollen nicht nur von *Statistik hören* sondern vor allem auch *Statistik sehen*, um so letztendlich *Statistik besser zu verstehen*.

Es galt deshalb, einen vorgetragenen mathematischen Satz oder Sachverhalt stets auch visuell aufzubereiten oder zumindest anhand eines nachvollziehbaren Beispiels im – den Vortrag begleitenden – *virtuellen Statistikbuch* zu erklären. Grafiken wie Box-Plots, QQ-Plots, u.ä. wurden ohnehin Schritt für Schritt ‘visuell hergeleitet’, aber auch bei der Erklärung fundamentaler Sätze – wie beispielsweise des zentralen Grenzwertsatzes – wurde versucht, durch schrittweise Visualisierung von Simulationsergebnissen das Verständnis vom *Kern der Sache* zu erleichtern.

3.4 Robustifizierung von fast ‘Allem’

Wesentliches Ziel war es, neben der *klassischen Statistik* von Beginn an in ganz natürlicher Weise auch die Grundzüge der *Robusten Statistik* zu vermitteln. Darauf wurde besonderer Wert gelegt, da gerade wirtschaftsstatistische Daten – wie sie unter anderem in der Statistik Austria vorkommen – stark mit ‘Ausreißern’ behaftet sein können. Zudem schien es sinnvoll, Kursteilnehmern mit bereits vorhandenen Kenntnissen nicht nur zur Auffrischung der beispielsweise im Zuge eines BWL- oder Soziologiestudiums gelernten klassischen Statistik zu dienen, sondern mit Überlegungen zu den Vorteilen robuster Methoden und Verfahren vielleicht auch neues Wissen anzubieten.

3.5 Aktive Teilnahme (fast) ‘Aller’

Eine Mischform aus Vortrag und Übung wird gerade in der Statistik/Mathematik als ideal erachtet. Die Kursteilnehmer sollten deshalb auch nicht stundenlang mit theoretischem Wissen berieselt werden, sondern von Beginn an aktiv am Geschehen teilnehmen. Die Durchführung zeitintensiver Übungen ist jedoch im vorliegenden Fall nicht möglich, insbesondere da es sich ja um Mitarbeiterschulungen handelt und ein Unternehmen naturgemäß nicht lange auf seine Mitarbeiter verzichten möchte.

Der Einsatz statistischer Softwaresysteme – wie beispielsweise **R**, **SAS**, **STATA** oder **SPSS** – zur Analyse einfacher Beispieldaten durch die Kursteilnehmer würde sicher zum Verständnis der zuvor vermittelten theoretischen Inhalte beitragen. Es müssten jedoch bestimmte Kenntnisse vorausgesetzt werden, wodurch bereits jener Teil der Interessierten vom Besuch der Grundlagenkurse ausgeschlossen wäre, der nicht so sehr mit der Handhabung spezieller statistischer Software vertraut ist. Aber auch für ‘Profis’ stellt sich die Frage, warum ein **R**-User für den Kursbesuch **SAS** verwenden sollte (oder umgekehrt), nur weil dieses Produkt in den praktischen Übungen eingesetzt wird.

Es galt also eine Lösung zu finden, wie unter den gegebenen Rahmenbedingungen alle Kursbesucher aktiv am Geschehen teilnehmen können. Fertige Applets und e-Learning Programme boten sich an, waren aber nicht zur Umsetzung aller Ideen und Vorstellungen

geeignet. Letztendlich wurde von den Autoren speziell für die computergestützte Lehre der interaktive *WissensCheck* in Form eines **R**-Packages mit Graphical User Interface (GUI) entwickelt (Templ and Dinges, 2007).

3.6 Interaktiver WissensCheck für ‘Alle’

Alle sollten (inter-)aktiv in den Kursablauf eingebunden werden. Das nach einem Vortrag obligatorisch in den Raum gestellte ‘*Gibt es zum Thema noch irgendwelche Fragen*’ soll schließlich nicht – wie üblich – von nur einigen wenigen beantwortet werden. Über ein eigens konzipiertes ‘TGUI’ sollte es auf einfache Weise möglich sein, gemeinsam ein vorgetragenes Thema mit einem speziell darauf zugeschnittenen *Wissens-Check* interaktiv abzuschließen.

Der WissensCheck

- **WAS** ist der *WissensCheck*: Training des zuvor behandelten Statistikstoffes durch kurze (Multiple-Choice-) Fragen oder kleine (graphische) Aufgaben.
- **WOZU** einen *WissensCheck*: durch aktive Teilnahme der Kursbesucher soll zuvor Vermitteltes gefestigt und der Kursablauf kurzweiliger gestaltet werden.
- **WIE** funktioniert der *WissensCheck*: einen Vortrag begleitend/ergänzend - interaktiv (und anonym) über das TGUI, einfach per Mausklick oder Slider.
- **VORTEIL** eines interaktiven *WissensCheck*: unmittelbares Feedback für beide Seiten (Kursleiter, Teilnehmer), keine langwierigen (zeitaufwendigen) Übungsabläufe.

Der *WissensCheck* ist (sportlich betrachtet) ein kleines Statistiktraining per Mausklick für zwischendurch. Ist im Anschluss an ein Thema der aufgerufene *WissensCheck* von allen absolviert worden, dann kann sofort gemeinsam die graphische Auswertung der Gruppenergebnisse betrachtet, und die zugehörige Lösung besprochen werden. Es ist dadurch auf einen Blick ersichtlich, ob der zuvor behandelte Stoff allgemein verstanden wurde, oder bestimmte Themen nochmals einer gezielten Erklärung bedürfen. Durch dieses begleitende *interaktive Gaming* werden die Kursteilnehmer von Beginn an aktiv eingebunden, die allgemeine Aufmerksamkeit erhöht und ganz ohne Prüfungsdruck (da Einzelergebnisse anonym bleiben) wird zuvor Vermitteltes gefestigt. Der *WissensCheck* soll Motivation, Lehr- und Lernunterstützung sein und sozusagen per Mausklick Feedback von ‘Allen’ für ‘Alle’ liefern.

4 Technische und praktische Umsetzung

Letztendlich galt es Ideen und Vorstellungen umzusetzen, um die (selbst-)gesteckten Ziele zu erreichen.

4.1 Verwendete Software

Zu allererst sollten die Kursbesucher schöne saubere Unterlagen vor Augen haben und dadurch *Statistik sehen*. Ideal geeignet für eine klare übersichtliche Darstellung der Inhalte ist ein Textsatzprogramm wie **L^AT_EX**, mit dem z.B. wissenschaftliche Texte verfasst werden können.

- **L^AT_EX** (www.latex-project.org) ist kein (*What You See Is What You Get*)-System, denn hier muss direkt der Quelltext verarbeitet werden. Zusätzlich kann (wie in unserem Fall) ein eigener ‘Style’-File geschrieben werden, der alle Vorstellungen und Wünsche berücksichtigt.

Zum besseren Verständnis mussten mathematische Inhalte visuell aufbereitet werden, d.h. es wurden schöne, ästhetische und ansprechende Grafiken benötigt, um auch komplexe Zusammenhänge anschaulich erklären zu können. Ideal geeignet dafür ist ein statistisches Programmsystem wie **R**.

- **R** (R Development Core Team, 2007), eine frei verfügbare open-source Software, ist ein System für statistische Analysen und Grafiken.

Einfache, nachvollziehbare Beispiele und Demos sollten die Inhalte zusätzlich verdeutlichen. Das bedeutete in der praktischen Umsetzung die Erzeugung von hunderten qualitativ hochwertigen Grafiken. Flexibilität war erforderlich um ein gutes Management von Folien und Grafiken zu ermöglichen. Unter Verwendung von **L^AT_EX** und **R** kann mit Hilfe von **Sweave** ein Dokument, welches bereits den Text sowie den Code für verschiedene Grafiken und Rechenoperation inkludiert, kompiliert werden. Änderungen an Grafiken u.ä. können dadurch direkt im ‘Text’ durchgeführt werden und unnötige Arbeitsschritte entfallen.

- **Sweave** (Leisch, 2002a, 2002b) zur dynamischen Verlinkung von Text und **R**-Code.

Zur interaktiven Visualisierung mehrdimensionaler Daten und zur Darstellung von Zusammenhängen (‘Highlighting’) durch verlinkte Grafiken wurde **JGR** eingesetzt.

- **JGR** (Helbig and Theus, 2005), *Jaguar* ist ein Java GUI (Graphical User Interface) für **R** und eine plattformunabhängige Lösung.

Animationen (‘Filmen’ in **R**), beispielsweise zur räumlichen Darstellung von Daten und als motivierende Beispiele mit ‘Aha-Effekt’, konnten durch **R** in Verbindung mit **ImageMagick** durchgeführt werden.

- **ImageMagick** (<http://www.imagemagick.org>).

Die Durchführung praktischer Übungen erfolgt über das **TGUI**, mittels interaktiven **WissensCheck** (Templ and Dinges, 2007). Der **WissensCheck** wurde in **R** mit Hilfe der Software **cttk** programmiert und kann unabhängig von individuellen Softwarekenntnissen der Kursteilnehmer eingesetzt werden. **cttk** wird auch für den bereits bekannten, ebenfalls für Schulungszwecke entwickelten **R-Commander** ((Fox, 2007)) eingesetzt.

- **cttk** (R Development Core Team, 2007) zur Erstellung von GUI’s in **R**.

Für das Einbinden von **R**-Grafiken in **cttk** Fenster wurde die Funktionalität des **R**-Packages **tkrplot** (Tierney, 2007) verwendet. Das Konzept zur Erstellung von *Slidern* wurde vom **R**-Package *TeachingDemos* (Snow, 2005) übernommen.

4.2 Beispiele aus dem virtuellen Statistikbuch

Grundidee war es, die Kursbesucher nicht einfach nur mit Wissen zu ‘berieseln’, sondern von Beginn an aktiv in das Geschehen mit einzubinden. Zum einen durch das Absolvieren praktischer interaktiver Übungen in Anschluss an einen Vortrag und zum anderen schon während des Vortrages selbst, indem von den Kursteilnehmern erstellte Daten bereits direkt in diverse Darstellungen der betrachteten Vortragsfolien eingebettet sind.

[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

Matrix-Schreibweise:

n Beobachtungen (Zeilen) und p Variablen (Spalten)

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

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)

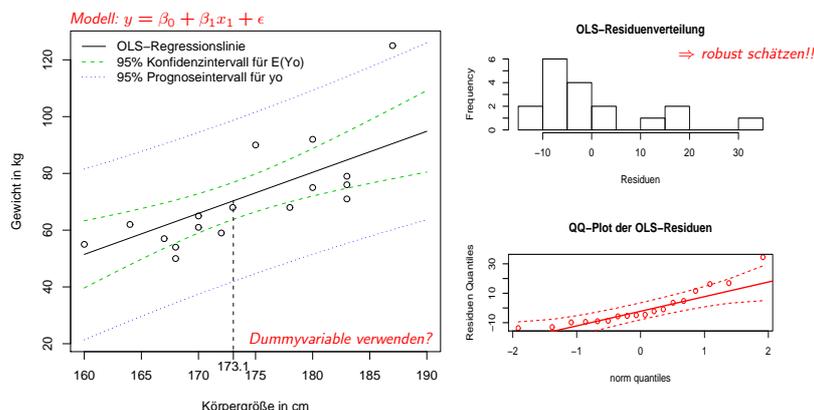
Quelle: ST.AT, 15. Mai 2005

Abbildung 1: Auszug aus dem virtuellen Statistikbuch, welches Schritt für Schritt einen Vortrag begleitet – selbst erstellte Daten der Kursteilnehmer, eingebunden als Matrix.

[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

Abbildung 2: Auszug aus dem virtuellen Statistikbuch, welches Schritt für Schritt einen Vortrag begleitet – selbst erstellte Daten der Kursteilnehmer, eingebunden in Grafiken.

Um die Daten der Kursteilnehmer in die Vortragsfolien einzubinden, greift der **R**-Code im **LaTeX** File auf die zuvor über ein kleines GUI (kurze Befragung unter einem Pseudonym) eingegebenen Daten zu und mit Hilfe von **Sweave** werden die Vortragsfolien mit aktualisierten Grafiken und Auswertungen in Sekundenschnelle neu erstellt.

Abbildung 1 zeigt jene Datenmatrix, die unmittelbar nach der Dateneingabe präsentiert wird. In Abbildung 2 werden die Daten zum Thema ‘*OLS-Diagnostik*’ anhand verschiedener Grafiken visualisiert. Bereits eingebunden in das virtuelle Statistikbuch können nun in weiterer Folge die eigenen Daten am Bildschirm in verschiedenen begleitenden Beispielen und graphischen Darstellungen betrachtet werden. Die besprochenen theoretischen Inhalte und Formeln werden dadurch – neben zahlreichen bereits fest eingebundenen Darstellungen – auch Schritt für Schritt durch Beispiele begleitet und veranschaulicht, die auf individuell erstellten Daten der Kursteilnehmer beruhen.

4.3 Beispiele aus dem WissensCheck

Ist ein vorgetragenes Kapitel abgeschlossen, dann erscheint am Bildschirm der Kursteilnehmer ein ‘*Alles klar?!*’. Dies ist die Aufforderung zur Teilnahme am *WissensCheck*.

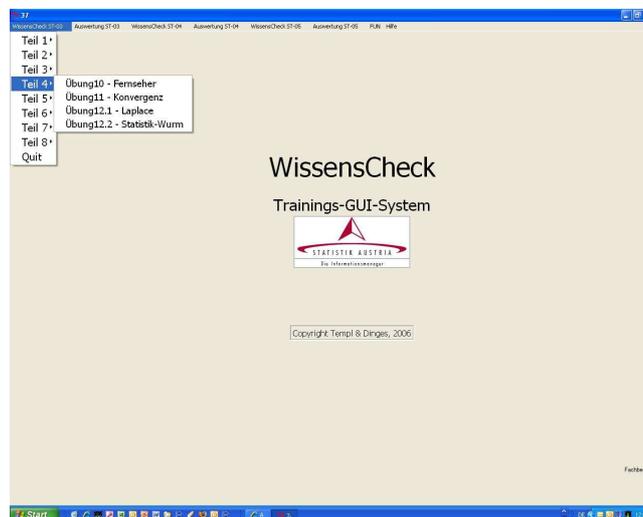


Abbildung 3: TGUI-Menü des *WissensCheck*.

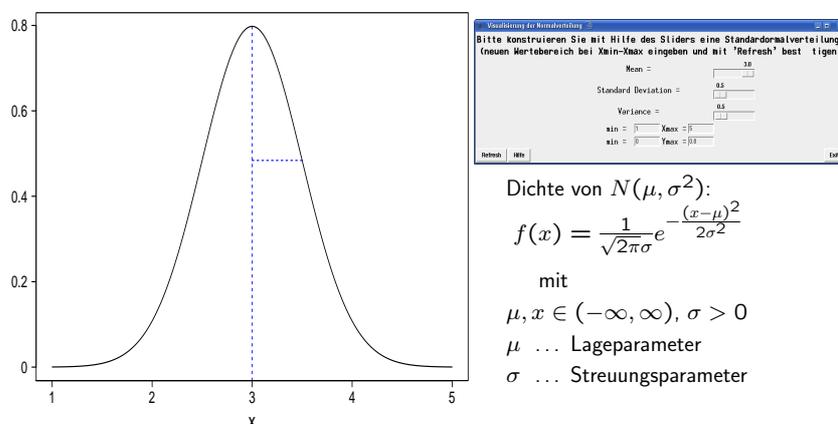


Abbildung 4: *WissensCheck*-Beispiel mit Slider – Thema ‘*Normalverteilung*’.

Über das Menü des TGUI (Abb. 3) können nun verschiedene Beispiele zum jeweiligen Thema absolviert werden. Entweder mit Hilfe eines Sliders, beispielsweise zur Simulation graphischer Darstellungen wie in Abbildung 4 dargestellt, oder auch einfach per Mausklick direkt in die Grafik bzw. durch die Beantwortung von Multiple-Choice-Fragen.

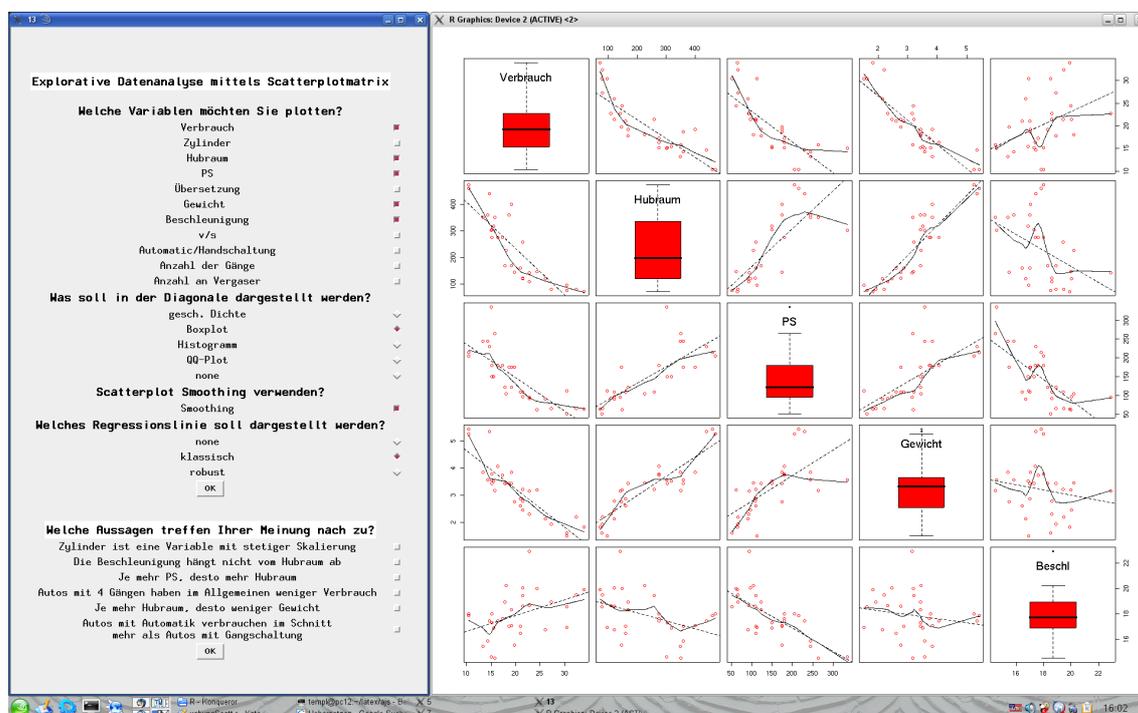


Abbildung 5: WissensCheck-Beispiel für Multiple-Choice-Fragen – Thema ‘Explorative Datenanalyse’.

In Abbildung 5 wird ein Beispiel des WissensCheck zum Thema ‘Explorative Datenanalyse’ veranschaulicht. Die (interaktive) Scatterplotmatrix kann anhand verschiedenster grafischer Tools individuell erstellt werden und ist explorativ zu nutzen, um die zugehörigen Multiple-Choice-Fragen zu beantworten. In Abbildung 6 wird demonstriert, wie wesentliche Inhalte auch spielerisch vermittelt werden können.

Unmittelbar nach Absolvierung der Übungen durch die Kursteilnehmer, werden die Ergebnisse gemeinsam besprochen. Die Beispiele werden dafür nochmals vom Kursleiter aufgerufen, nun jedoch bereits mit Lösung wie in der linken Hälfte von Abbildung 7 dargestellt. Rechts davon scheint nun die Antwortverteilung aller Kursteilnehmer auf, wodurch auf einen Blick ersichtlich ist, ob das Beispiel allgemein verstanden wurde oder nochmals einer gezielten Erklärung bedarf.

Der WissensCheck dient jedoch nicht nur als Abschluss und zur Vertiefung eines zuvor behandelten Kapitels, sondern auch zur Vorbereitung und Überleitung auf ein nachfolgendes Thema. In Abbildung 8 wird demonstriert, wie mit einem speziellen Beispiel vom Kapitel ‘Einfache lineare Regression (OLS)’ auf die nachfolgenden Themen ‘Einfluss von Ausreißern’ und ‘Robuste Regressionsverfahren’ übergeleitet wird.

Abbildung 9 zeigt nach absolvierter Übung die Auflösung des Beispiels durch den

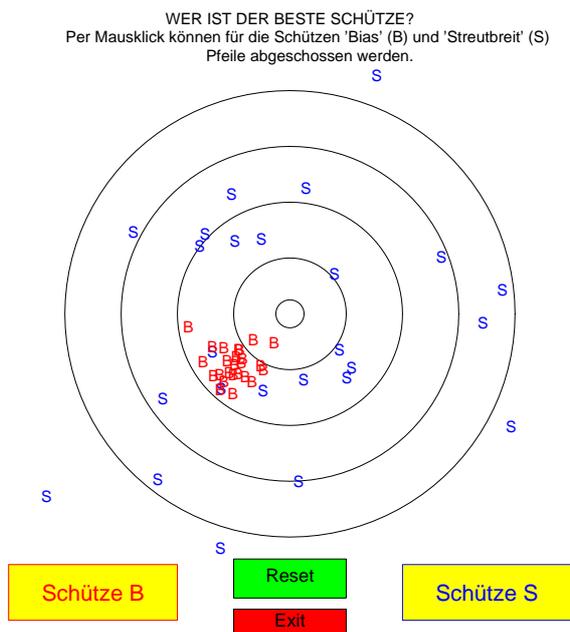


Abbildung 6: WissensCheck-Beispiel für Spielerisch Lernen – Thema ‘Streuung & Bias’.

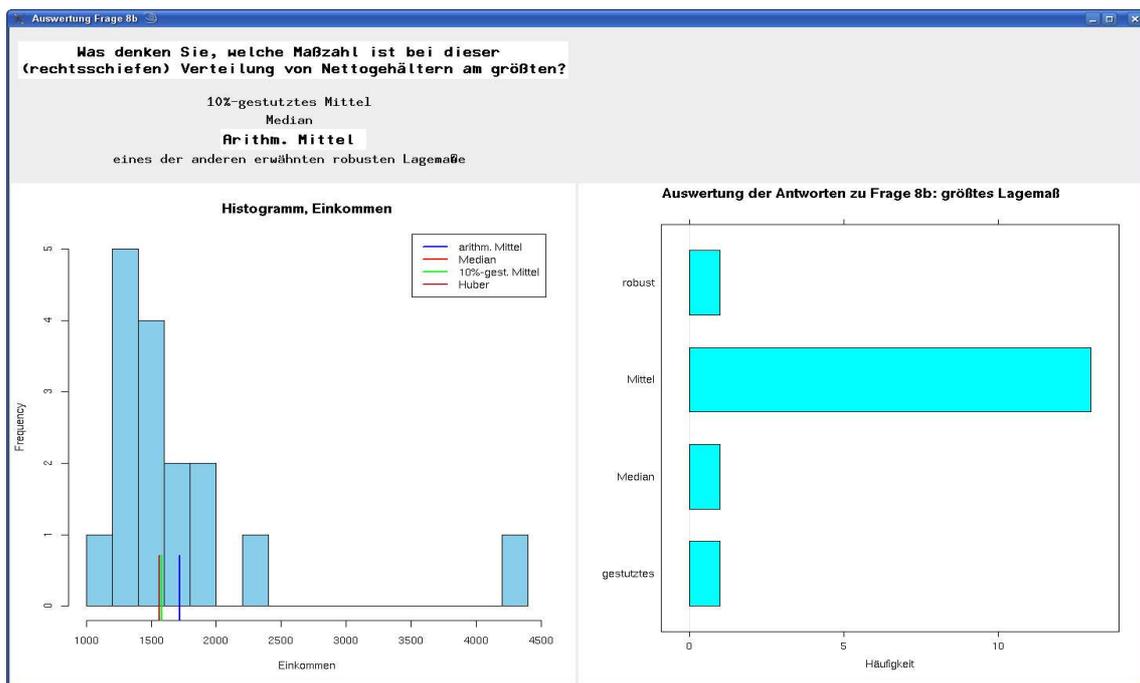


Abbildung 7: Absolviertes Beispiel des WissensCheck, mit Lösung und Antwortverteilung der Kursbesucher – Abschluss eines zuvor behandelten Themas.

Kursleiter. In der linken Hälfte der Abbildung ist wieder die zuvor gestellte Frage mit eingezeichneter Lösung dargestellt. Die Kursteilnehmer konnten bei der Beantwortung der Frage natürlich nur die ‘nackte’ Punktwolke, ohne die hier bereits eingezeichneten

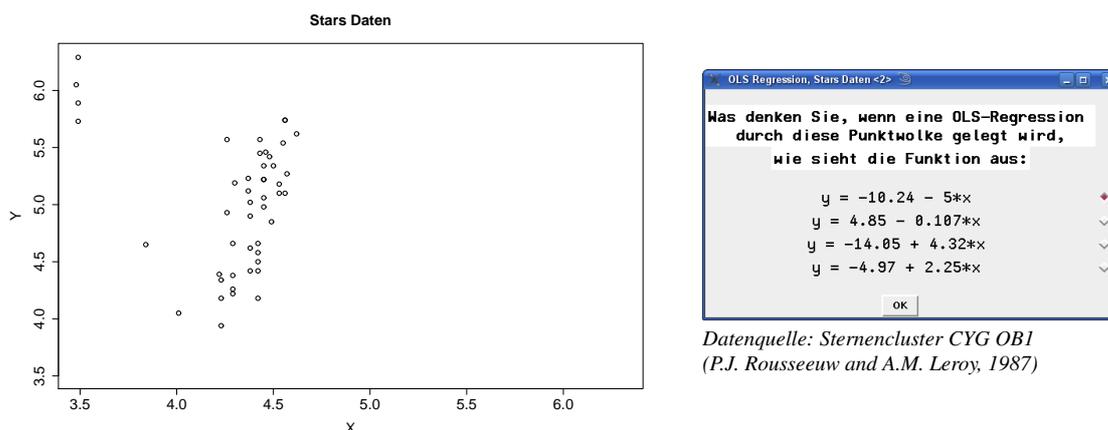


Abbildung 8: Beispiel des WissensCheck – Vorbereitung auf ein nachfolgendes Thema.

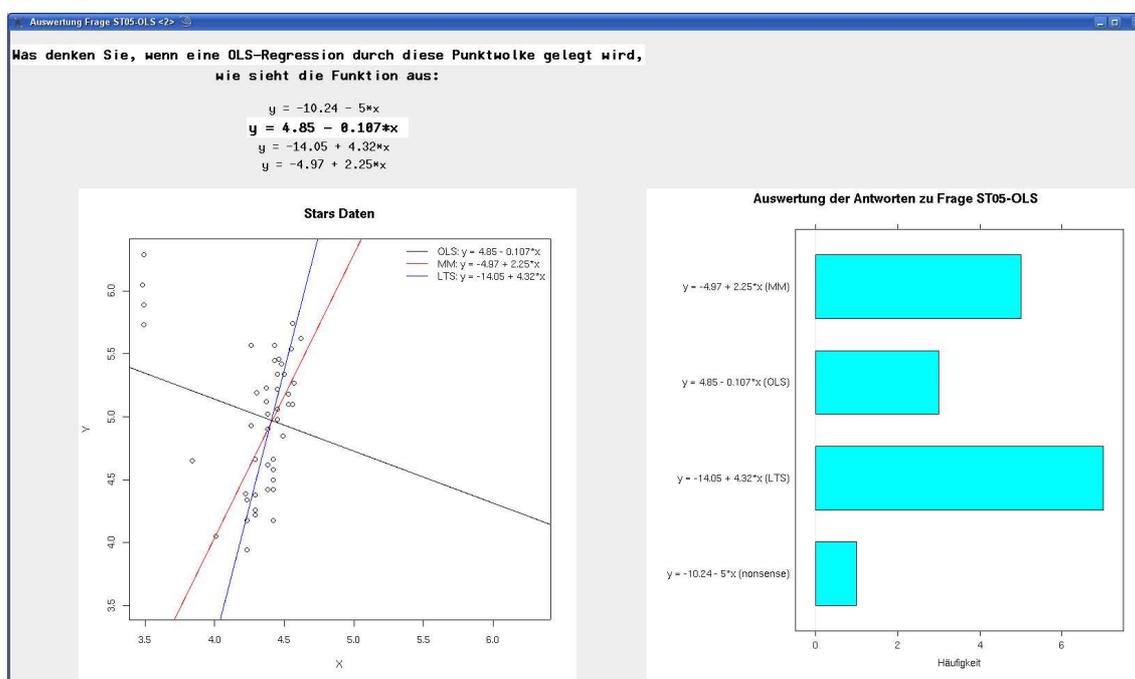


Abbildung 9: Absolviertes Beispiel des WissensCheck, mit Lösung und Antwortverteilung der Kursbesucher – Vorbereitung auf ein nachfolgendes Thema.

Regressionslinien und richtig markierter Gleichung betrachten. Ihre Antwortverteilung³ auf der rechten Seite der Abbildung ist nun nicht mehr so eindeutig wie noch zuvor in Abbildung 7, wo es sich ja um die Wiederholung eines bereits vorgetragenen Themas handelte. Eine mit dem bisher Gelernten scheinbar leicht lösbare Aufgabe, stellt sich nun für manchen als ‘*statistische Falle*’ heraus. Dies kann in gleicher Situation auch in der

³Die in Abb. 7 und 9 dargestellten Antwortverteilungen dienen zur Demonstration der Kursgestaltung und basieren nicht auf erhobenen Echtdate. Die tatsächlichen WissensCheck-Auswertungen einer Gruppe werden nur während des Kurses, und ausschließlich den beteiligten Kursteilnehmern präsentiert!

Praxis passieren, wo immer noch häufig keine Visualisierung erfolgt, und sollte den Teilnehmern bereits im Vorfeld die Sinnhaftigkeit des nachfolgenden Kapitels verdeutlichen und zur Vorsicht bei der Anwendung nicht robuster Regressionsverfahren mahnen.

Die in diesem Artikel präsentierten Beispiele stellen einen kleinen Ausschnitt aus *virtuellem Statistikbuch* und *WissensCheck* dar. Das entwickelte **R**-Package (TGUI-Version 5.2) enthält derzeit über 500 grafische Darstellungen sowie unzählige Tabellen und Formeln. Etwa 430 dieser Grafiken sind in Beispiele des *virtuellen Statistikbuches* eingebunden, wobei ein Teil davon direkt auf von den Kursbesuchern selbst erstellte Daten zugreift und somit stets individuell erzeugt wird. An die 80 (interaktive) Beispiele sind überdies in der derzeitigen *WissensCheck*-Version 5.2 enthalten.

5 Zusammenfassung und Ausblick

Zu Beginn schien es eher schwierig alle Ideen und Vorstellungen umzusetzen und die (selbst-)gesteckten Ziele zu erreichen. Die umfangreichen Inhalte sollten aufgrund des engen Zeitrahmens in sehr kompakter Form behandelt werden und dennoch das Wesentliche vermitteln. Es sollte darüber hinaus gelingen, alle Beteiligten auf eher ungezwungene und kurzweilige Art und Weise zur aktiven Kursteilnahme zu ‘motivieren’.

Zum einen waren dafür abstrakte mathematische Inhalte und Theorien durch Visualisierung möglichst verständlich und nachvollziehbar aufzubereiten – um so der oft gefürchteten ‘trockenen’ Statistik etwas den Schrecken zu nehmen. Zum anderen sollte gemäß dem Motto ‘*take the right tool*’ gezeigt werden, dass hinter vermeintlich simplen Berechnungen oder Grafiken auch so manch (statistische) Falle lauern kann. Es sollte bewusst gemacht werden, dass nicht hinter jeder beliebigen Zahl *die Statistik* steckt, und insbesondere, dass nicht jeder der weiß, dass der Median kein Schweizer Gebirge sein kann, sogleich *ein Statistikexperte* sein muss. Die oftmals falsche Wahrnehmung der Statistik in der Öffentlichkeit lässt ohnehin den einen oder anderen von uns frühzeitig ergrauen – man denke nur an das ‘*und nun für die Statistiker unter uns*’ bei einer Fußballübertragung; Statistik ist schließlich kein ‘Erbsenzählen’, sondern eine hochentwickelte *Wissenschaft*. Dieses Wissen, das durch die Statistik entsteht, trägt unter anderem zu Fortschritt in Medizin und Technik bei, dient zur Entscheidungsfindung in Wirtschafts- und Sozialpolitik und liefert den Bürgern seriöse Informationen zu Wirtschaft, Politik und Gesellschaft. Jedem sollte es deshalb ein Anliegen sein statistisches Basiswissen zu erlangen – nicht zuletzt auch um manch (un)bewusste mediale Irreführung leichter zu durchschauen. ‘Eine grundlegende statistische Ausbildung sollte zur Allgemeinbildung jedes mündigen Staatsbürgers gehören’⁴.

Dies alles war Anlass und Antrieb, das hier präsentierte Lernsystem in der nun vorliegenden Form zu konzipieren und so grundsätzlich jedem Interessierten, unabhängig von individuellen Software- oder Fachkenntnissen, eine Kursteilnahme zu ermöglichen – (inter-)aktiv, ‘anonym’ und ohne Prüfungsdruck⁵. Das entwickelte System soll Motivation, Lehr- und Lernunterstützung sein und – einfach per Mausclick – unmittelbares Feedback von ‘Allen’ für ‘Alle’ liefern. Bisherige Erfahrungen aus den seit 2005 statt-

⁴Vgl. <http://www.stat4u.at>, ein Projekt des Instituts für Scientific Computing der Universität Wien.

⁵Vgl. http://www.osg.or.at/download/news/39/15%20Dinges_Temp1.pdf

findenden Schulungen zeigen, dass – nach meist kurzer anfänglicher Skepsis – insbesondere der WissensCheck rasch auf allgemeine Begeisterung stößt, und sich durch dessen Interaktivität der gesamte Kursablauf für Teilnehmer und Vortragende interessanter und kurzweiliger gestaltet.

Das zur Lehr- und Lernunterstützung konzipierte **R**-Package (TGUI-Version 5.2) mit *virtuellem Statistikbuch* und *interaktivem WissensCheck* stellt ein erstes Basissystem dar, welches laufend modifiziert und erweitert wird; neue (Er-)Kenntnisse, Erfahrungen und Ideen werden auch künftig einfließen. Der Einsatz ist unabhängig vom Betriebssystem möglich und - mit einem speziell zugeschnittenen WissensCheck – auch zur praktischen Unterstützung anderer Fachbereiche denkbar.

Vorrangiges Ziel der eigenen Entwicklungsarbeiten war es, ein modernes und effektives Instrument zu gewinnen und so die methodischen Grundlagenkurse, die zur laufenden Mitarbeiterschulung in der Statistik Austria angeboten werden, zu unterstützen. Ein externer Einsatz des Trainingstools war ursprünglich nicht geplant und zusätzliche computergestützte Schulungen außerhalb von Statistik Austria nicht vorgesehen. Es könnte jedoch angedacht werden – externer Bedarf und erforderliche interne Ressourcen vorausgesetzt – das entwickelte TGUI-Package hinsichtlich der unterschiedlichen und neuen Anforderungen zu adaptieren, und dieses als standardisiertes/aufbauendes computergestütztes Statistik-Trainingstool auch anderen Institutionen, Unternehmen oder Schulen zugänglich zu machen. Vielleicht könnte damit ja ein kleiner Beitrag dazu geleistet werden, mitunter auch den einen oder anderen ‘Statistik-Skeptiker’ zu bekehren.

Derzeit wird das beschriebene lehr- und lernunterstützende Statistiktraining von den STAT-Methodikern ausschließlich zur hausinternen Mitarbeiterschulung in der Statistik Austria angeboten; durch den hier präsentierten Einblick sollte es jedoch auch anderen Lehrenden und Institutionen als konzeptioneller Anreiz oder alternative Idee einer effektiven Kursgestaltung vorgestellt werden.

Danksagung

Ein *merci* an die Verantwortlichen der Statistik Austria, die den Einsatz neuer wissenschaftlicher Methoden und moderner Verfahren stets unterstützen, und uns auch bei dieser Projektumsetzung wieder vertrauensvoll freie Hand gaben. Insbesondere den Teilnehmern der ersten Methodikkurse gebührt jedoch Dank für Kritik und Verbesserungsvorschläge, die wesentlich zur Kursmodifizierung beitrugen, und für das positive Feedback, welches erst die nötige Motivation gab, die oftmals sehr arbeitsintensiven Weiterentwicklungen nicht nur im Job in Angriff zu nehmen.

Literatur

- Fox, J. (2007). Rcmdr: R Commander [Computer software manual]. Available from <http://www.r-project.org>, <http://socserv.socsci.mcmaster.ca/~jfox/Misc/Rcmdr/> (R package version 1.3-5)
- Helbig, M., and Theus, M. (2005). JGR: Java GUI for R. *Statistical Computing & Graphics*, 16, 8-12.

- Leisch, F. (2002a). Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle and B. Rönz (Eds.), *Compstat 2002 – Proceedings in Computational Statistics* (p. 575-580). Heidelberg: Physica Verlag. Available from <http://www.stat.uni-muenchen.de/~leisch/Sweave>
- Leisch, F. (2002b). Sweave, part I: Mixing R and L^AT_EX. *R News*, 2, 28-31. Available from <http://CRAN.R-project.org/doc/Rnews/>
- R Development Core Team. (2007). *R: A Language and Environment for Statistical Computing [Computer software manual]*. Vienna, Austria. Available from <http://www.R-project.org>
- Snow, G. (2005). *TeachingDemos: Demonstrations for teaching and learning [Computer software manual]*. (R package version 1.5)
- Templ, M., and Dinges, G. (2007). *TGUI: Trainings GUI [Computer software manual]*. (R package version 5.2, zur internen Nutzung in der Statistik Austria / Copyright Statistik Austria)
- Tierney, L. (2007). *tkrplot: TK Rplot [Computer software manual]*. (R package version 0.0-16)

Adresse der Autoren:

Gerlinde Dinges
Fachbereich Methodik
Abteilung Register, Klassifikation und Methodik
Statistik Austria
Guglgasse 13
A-1110 Wien

E-Mail: gerlinde.dinges@statistik.gv.at
<http://www.statistik.at/>

Matthias Templ
Institut für Statistik und Wahrscheinlichkeitstheorie
Technische Universität Wien
Wiedner Hauptstraße 8-10
A-1040 Wien

E-Mail: templ@statistik.tuwien.ac.at
<http://www.statistik.tuwien.ac.at/public/templ/>