Über die Möglichkeiten des Einsatzes von R in der Statistikausbildung

Bernhard Spangl,

Universität für Bodenkultur, Wien

Bad Doberan, 1. Okt. 2009

Einsatz von R . . .

- . . . in der Lehre
  - als Statistikprogramm
  - zur Lösung von Übungsaufgaben
- . . . zur Generierung von
  - Übungsaufgaben
  - Tests
  - Prüfungen

Vorbemerkungen

- viele Studierende ($\sim$700 im WS, $\sim$250 im SS)
- früh im Studienplan (1., 2. oder 3. Semester), abhängig von der Studienrichtung
- hauptsächlich Windows-Benutzer
- nicht vertraut mit ‘open source’ oder ‘mirrors’ (oft sind es auch die Master-Studierenden nicht!)
- keine weiteren Ressourcen (zeitlich, personell, EDV-Räume, etc.) vorhanden, um den Studierenden eine zusätzlichen Einführung in R zu geben
  $\hookrightarrow$ wir stellen nur eine modifizierte Version von R zur Verfügung
  + Skriptum

Modifizierungen

- Herunterladen & Installieren von R-2.9.2-win32.exe
- benutzerdefinierte Installation, lediglich Auswahl von
  - Main Files
  - Compiled HTML Help Files
  - Support Files for Package tcltk
  - Message Translation
- Installation zusätzl. Pakete, z.B. foreign, xlsReadWrite
- Erzeugen des Verzeichnises ‘eigene R-Dateien’
- Erzeugen eines entsprechenden Batch-Programms, um R zu starten
- Erzeugen von `.First` & Default-Workspace ‘.RData’
- Erzeugen von ‘R-Ordner.zip’

Batch-Programm & `.First`

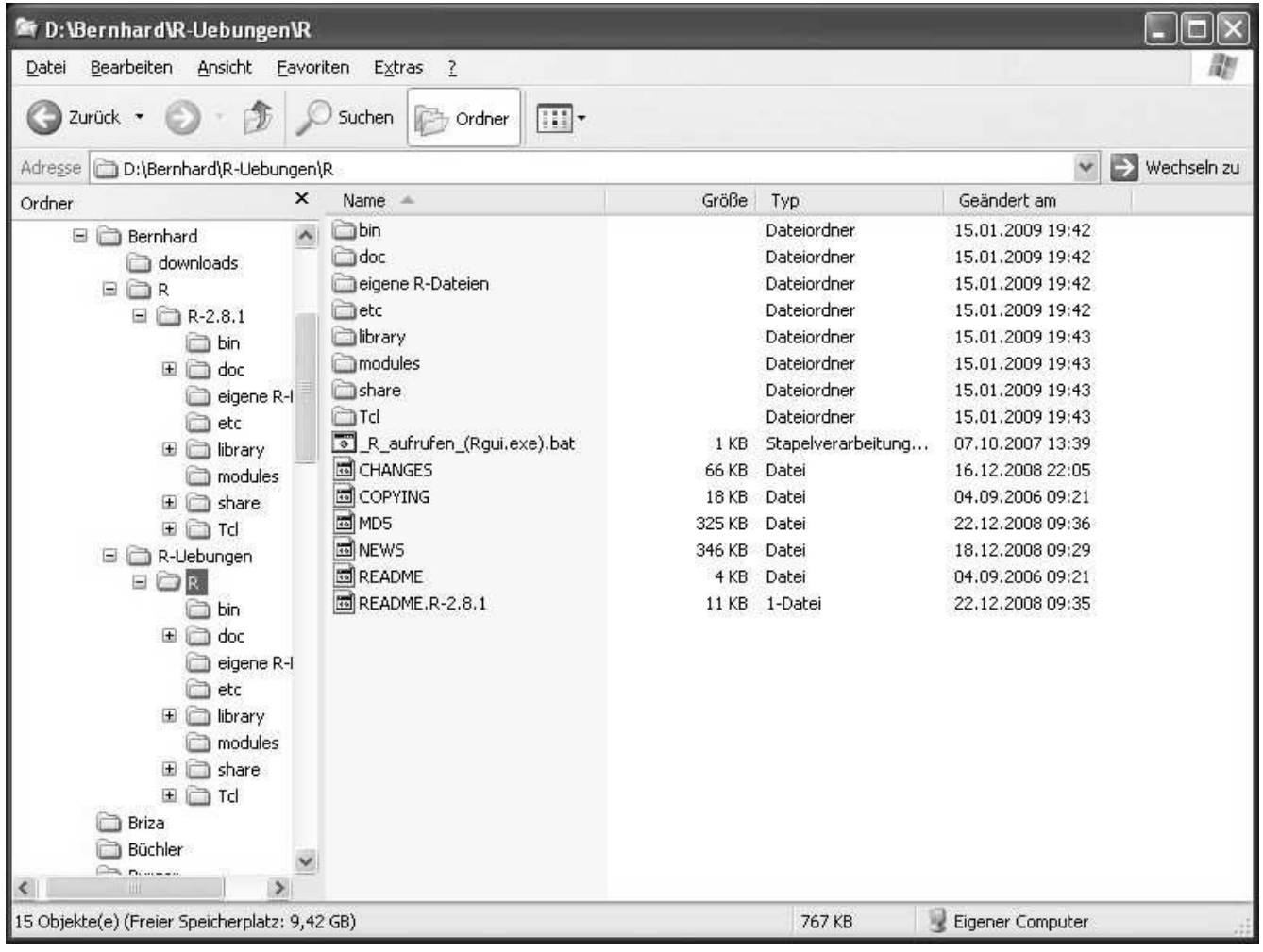
- Batch-Programm:

```
@echo off
echo Rgui wird gestartet ...
cd "eigene R-Dateien"
start ..\\bin\\Rgui.exe --sdi
```

- R-Funktion `.First`:

```
.First <- function () {
library(xlsReadWrite)
[...]
cat(getwd(), "\n")
[...]
}
```

Screenshot Explorer



ANOVA

Die entsprechenden Nullhypothesen für die Varianzanalyse lauten:
H_{Sorte}: Die Sorte hat keinen Einfluss auf den theoretischen mittleren Ertrag.
$H_{Steinmehl}$: Die Zugabe von Steinmehl zur Gülle hat keinen Einfluss auf den theoretischen mittleren Ertrag.
$H_{Steinmehl:Sorte}$: Es gibt keine Wechselwirkungen zwischen den beiden Faktoren Sorte und Steinmehl.
Die Alternativhypothesen lauten: "Sorte und Zugabe von Steinmehl zur Gülle haben Einfluss auf den theoretischen mittleren Ertrag und es gibt Wechselwirkungen zwischen den beiden Faktoren." Das Risiko 1. Art wird mit α =0.05 festgelegt.

```
> anova(lm(roggen$ertrag ~ roggen$steinmehl * roggen$sorte))
```

'+' ..... Modell ohne Wechselwirkungen
'*' ..... Modell mit Wechselwirkungen
'*' steht kurz für das Modell 'sorte + steinmehl + sorte: steinmehl'

```
Analysis of Variance Table

Response: roggen$ertrag
                           Df  Sum Sq Mean Sq F value    Pr(>F)
roggen$steinmehl            1  11.900  11.900  7.3731  0.014181 *
roggen$sorte                2  73.091  36.545 22.6424 1.218e-05 ***
roggen$steinmehl:roggen$sorte  2  25.556  12.778  7.9168  0.003414 **
Residuals                  18  29.052   1.614
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Die p-Werte für alle drei Hypothesen liegen unter dem gewählten Risiko 1. Art α = 0.05, daher müssen alle Hypothesen verworfen werden. Es haben daher sowohl die Zugabe von Steinmehl zur Gülle als auch die Sorte Einfluss auf den Ertrag und es gibt Wechselwirkungen zwischen diesen beiden Faktoren.
Die Mittlere Quadratsumme des Fehlers (1.614) entspricht der Fehlervarianz s_ε^2.

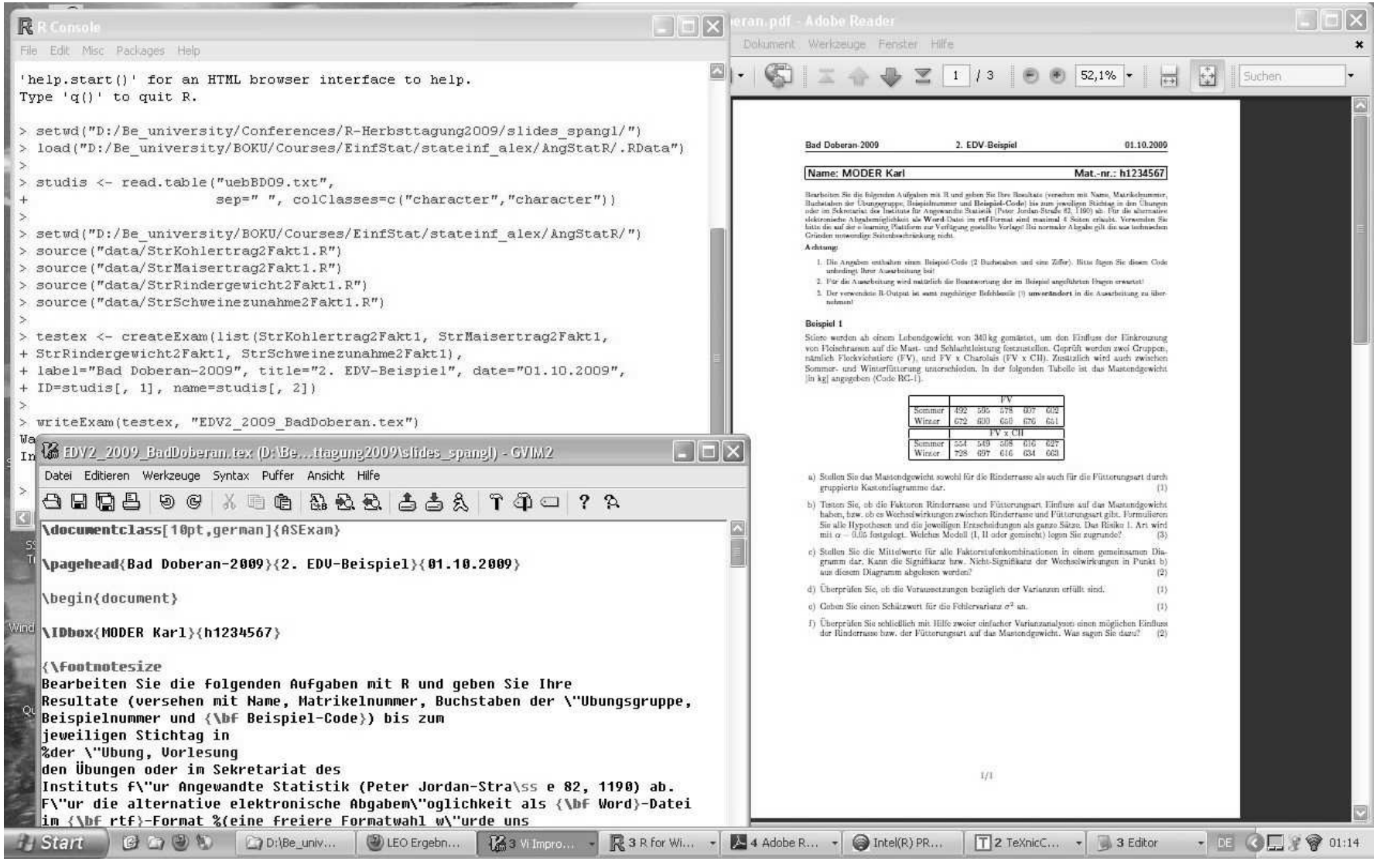
Anmerkungen

- Vorteile:
  - Kombination der modifizierten R-Version + Skriptum gut von den Studierenden angenommen
  - portabel, einfach auf USB-Stick kopieren
  - aus didaktischer Sicht:
    Die Studierenden hinterfragen ihre Resultate.
    (Haben sie nie getan, als SPSS verwendet wurde!)
- Nachteil:
  - nur Anwenden von ‘Kochrezepten’, kein richtiges Verständnis/Erlernen von R

Übungsaufgaben

- Konzept: Alexander Ploner
- Verknüpfung von R und Latex
- beliebig viele, individuelle Übungsangaben
- seit etlichen Jahren im Einsatz
- umfangreiche Funktionalität breits vorhanden, musste allerdings erst geschaffen werden
- mittlerweile überholt $\longrightarrow$ Sweave

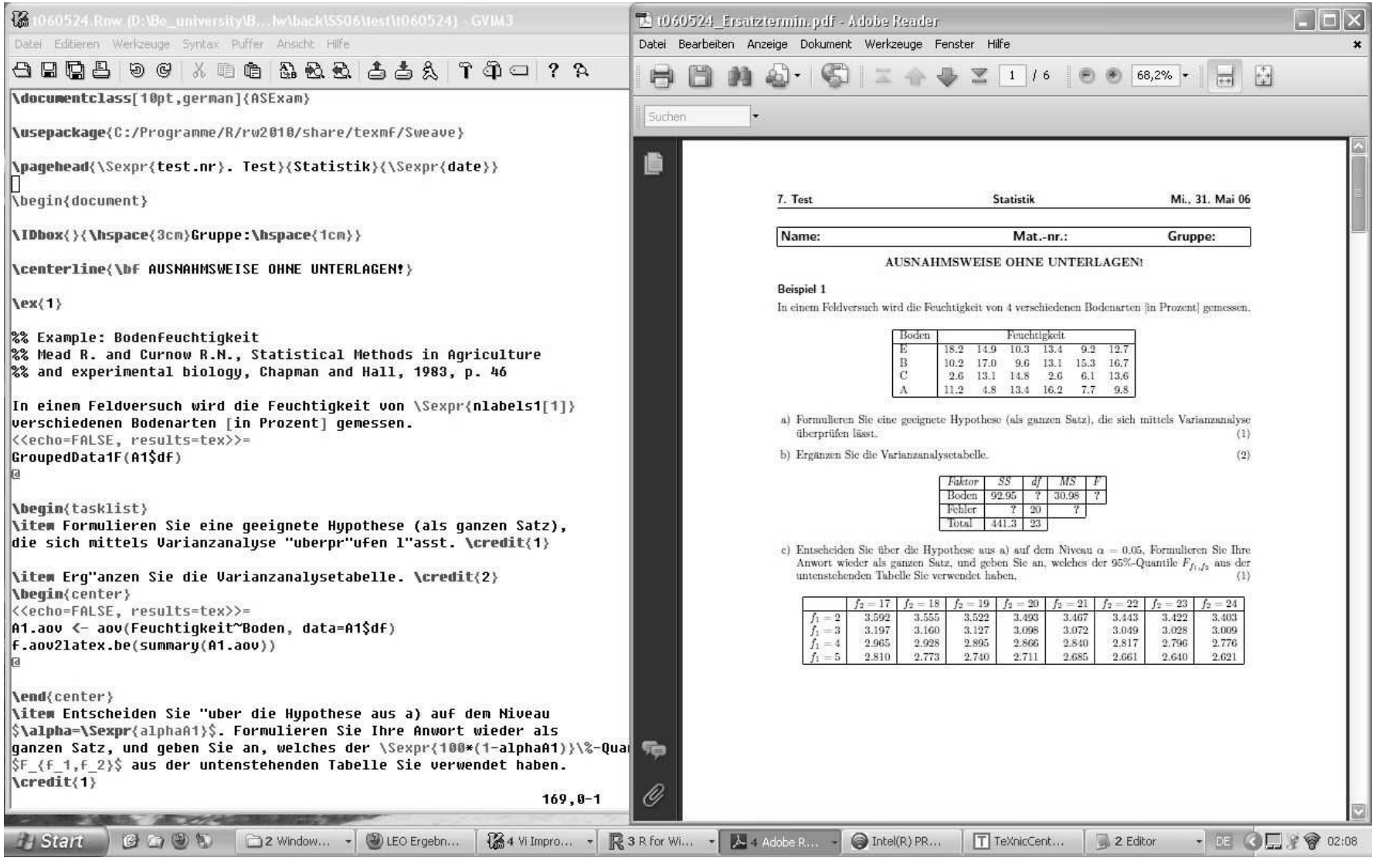
Screenshot Übungsaufgaben



Tests

- Verknüpfung von R und Latex mittels ‘Sweave’
- verschiedene, gleichwertige Testgruppen
- Übungsgruppen an verschieden Wochentagen
- aufbauend auf bereits vorhandener Funktionalität, plus eigene neue Funktionen
- mit Lösungen!

Screenshot Tests



Diskussion

- R ist kommandozeilen-orientiert (CLI)
- graphische Benutzeroberfläche (GUI)
  - R-Paket ‘Rcmdr’
  - jedoch nur eingeschränkte Funktionalität
- Alternativen für Tests/Prüfungen:
  - R-Paket ‘exams’
  - an WU Wien entwickelt und eingesetzt

Screenshot R-Commander