



Approaches

Using R for teaching statistics to nonmajors:
Comparing experiences of 2.5 different approaches

- Based on guided code generation
- Based on exploring existing code
- Based on spreadsheet interaction

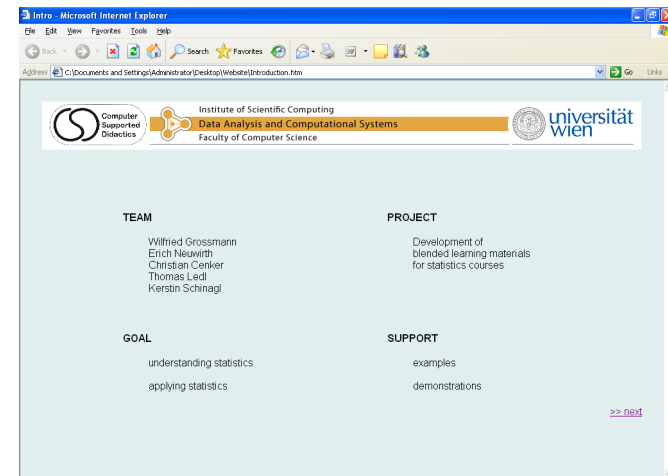
Thomas Baier
Richard Heiberger
Kerstin Schinagl
Erich Neuwirth

University of Technology, Vienna
University of Vienna
Temple University, Philadelphia

Approaches

- Based on guided code generation
 - RCommander and extensions
 - Temple University has a course based on this approach
- Based on exploring existing code
 - RPad
 - University of Vienna has a course based on this approach
- Based on spreadsheet interaction
 - RExcel
 - University of Vienna has lab classes based on this approach

RPad



RPad

Startseite LV - Microsoft Internet Explorer

Computer Supported Didactics Institute of Scientific Computing Data Analysis and Computational Systems universität wien Faculty of Computer Science


Statistics and Probabilities in Computer Sciences

CHAPTERS	MODULES			
Introduction to Probability	Basics	Demonstration		
Applications of Probability	Introduction	Demonstration	Terminology and Methodology	Analysis Methods
Descriptive Statistics	Introduction	Demonstration	Terminology and Methodology	Analysis Methods
Statistical Inference	Introduction	Demonstration	Terminology and Methodology	Analysis Methods
Analysis of Frequencies	Introduction	Demonstration	Terminology and Methodology	Analysis Methods
Analysis of Variance	Introduction	Demonstration	Terminology and Methodology	Analysis Methods
Regression Analysis	Introduction	Demonstration	Terminology and Methodology	Analysis Methods

RPad


Demo Flat Share

HOW ABOUT living in such a residence?



BUT HOW ABOUT COSTS ??
- UNAFFORDABLE

A very cheap alternative could be




BUT HOW ABOUT COMFORT ...

A popular and also relatively low priced form of living is the **flat share**. Especially students like this style of living. As the rental charges differ a lot we want to detect the main features responsible for favourable rents.

RPad

Demo Flat Share

HOW ABOUT living in such a residence?




BUT HOW ABOUT COSTS ??

A popular and also relatively low priced form of living is the **flat share**. Especially students like this style of living. As the rental

RPad

Demo Flat Share



A popular and also relatively low priced form of living is the **flat share**. Especially students like this style of living. As the rental charges differ a lot we want to detect the main features responsible for favourable rents.

DATASET

- Does the amount of rent depend on the size of the room?
- Are there differences in the rental charges due to the number of flat mates sharing an accommodation?

Analysis in R

Analysis in R (hiding code)

RPad

http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad - Microsoft Internet Explorer

Address: http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad

Demo Flat Share

The variable **rent** presents the dependent variable. The variables **size** and **number** are the independent variables also called factors. We want to find out if the number of flat mates (two, three, more than three) has an effect on the average costs of rent.

- Descriptive Statistics
- Model Analysis
- Model Assumptions
- Interpretation

Path: [home](#) > [h1](#)

RPad

http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad - Microsoft Internet Explorer

Address: http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad

Demo Flat Share

The variable **rent** presents the dependent variable. The variables **size** and **number** are the independent variables also called factors. We want to find out if the number of flat mates (two, three, more than three) has an effect on the average costs of rent.

- Descriptive Statistics
 - What's the average rent?
 - How much does the cheapest room cost?
 - Are there big differences in the monthly costs?
 - Is there a possibility to visualize the effect of the number of flat mates on the rent?
- Model Analysis
- Model Assumptions

Calculate

Path: [home](#) > [blocknote](#) > [q](#)

RPad

http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad - Microsoft Internet Explorer

Address: http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad

Summary Statistics

	mean	min	max	std.dev.	counts
Total	271.7	150	435	53.9	54
two	300.2	180	435	56.5	18
three	255.7	180	360	41.1	18
more	249.1	150	334	52.1	18

Boxplot Rent

Path: [home](#) > [table](#) > [h0](#) > [h1](#) > [h2](#) > [h3](#) > [fields](#) > [blocknote](#) > [q](#) > [span](#) > [rent](#)

RPad

http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad - Microsoft Internet Explorer

Address: http://127.0.0.1:8079/Wohngemeinschaften_OC.Rpad

Model Analysis

The Analysis of Variance is used to check an effect on the average rent due to the number of flat mates.

The null hypothesis and alternative hypothesis means:

H_0 : The number of flat mates has no effect on the average rent per person.

H_a : The number of flat mates has an effect on the average rent per person.

Calculate

ANOVA Table

Analysis of Variance Table

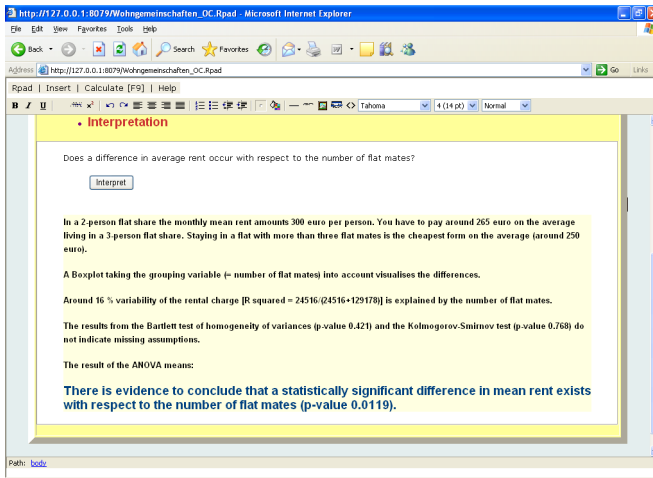
Response: rent	Df	Sum Sq	Mean Sq	F value	Pr(>F)
number	2	23616	11808	4.8386	0.0119
Residuals	51	129178	2533		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients

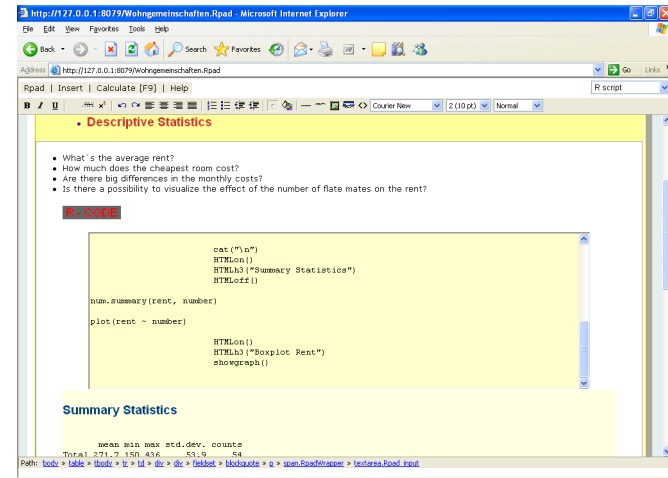
Path: [home](#) > [table](#) > [h0](#) > [h1](#) > [h2](#) > [h3](#) > [fields](#) > [blocknote](#) > [q](#) > [span](#) > [rent](#)

RPad

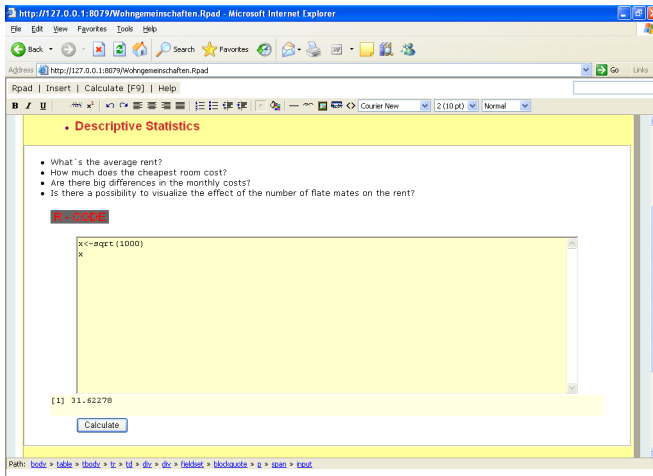


RPad

RPad



RExcel

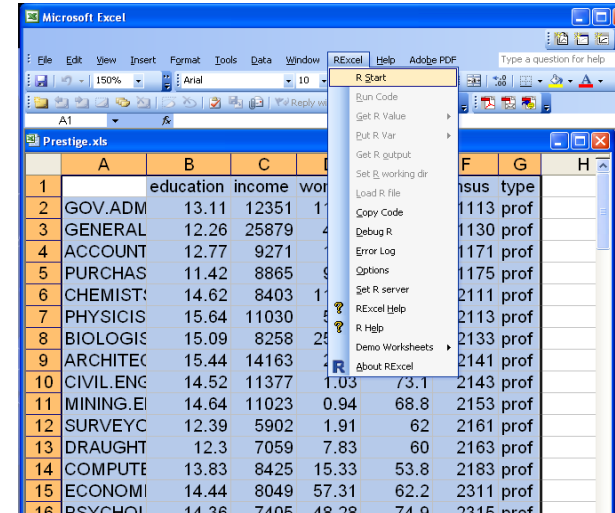


- Approach allows students to use a tool they already know and embed advanced statistical methods in the spreadsheet paradigm
- At the same time, prior spreadsheet experience can help with data manipulation and graphics creation

RExcel and RCommander

- Temple University currently teaches an introductory statistics course based on RCommander and RExcel.
- Statistics with Excel always needs an addin. We use R, the best possible addin.
- We use the RExcel interface to get data from Excel into R and to get tabular results back from R.
- RCommander provides a clickable menu interface to R.

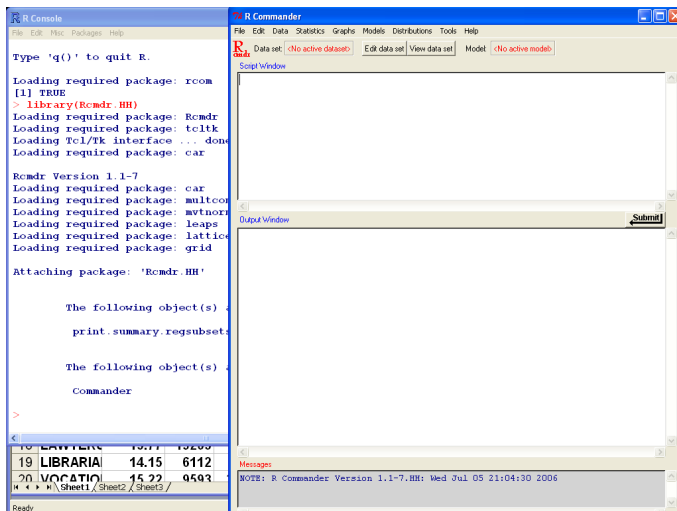
RExcel and RCommander



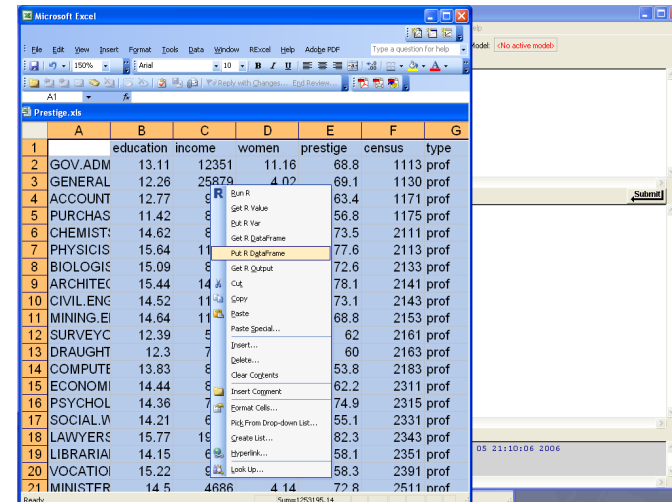
- Excel spreadsheet containing the Prestige dataset from the car package
- Click on the RExcel menu item to start R

RExcel and RCommander

RExcel and RCommander

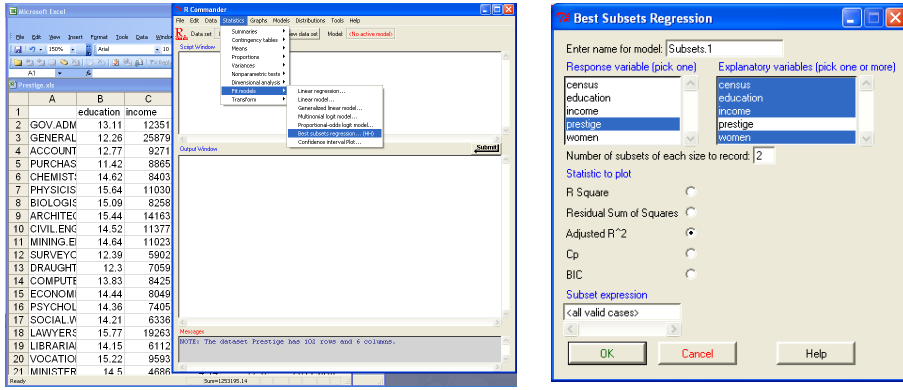


- In the R GUI, enter library(Rcmdr.HH)
- This starts the Rcmdr window with the HH menu



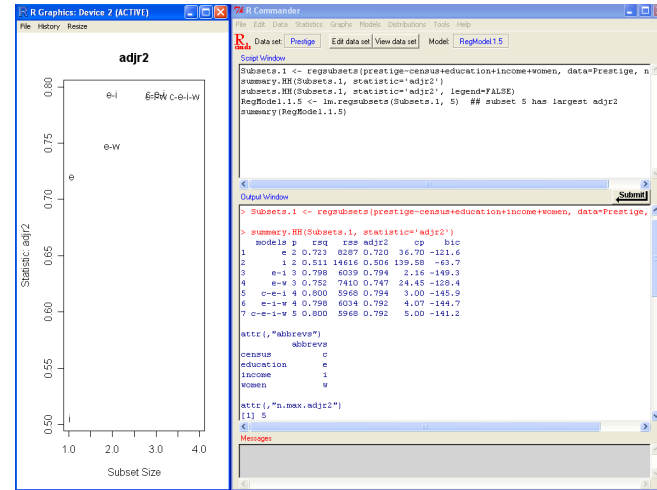
- Highlight and right click a region in Excel
- Send the region to Rcmdr

RExcel and RCommander



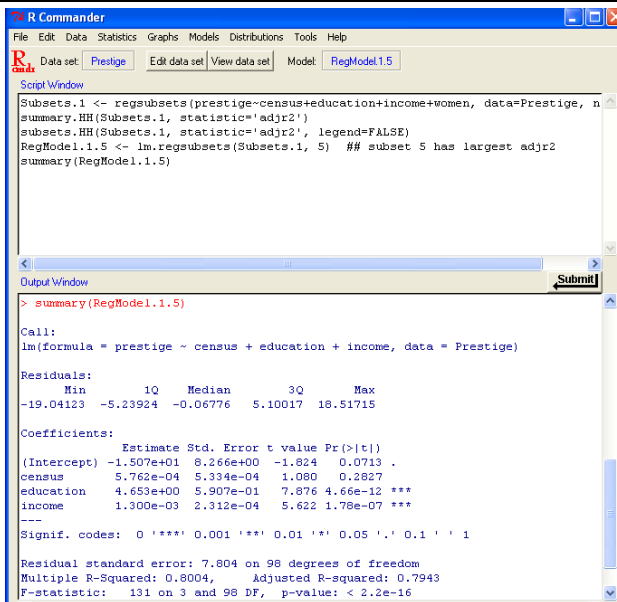
- Click the “Statistics/Fit Models/Best Subsets Regression” menu
- Fill in the model specification box

RExcel and RCommander



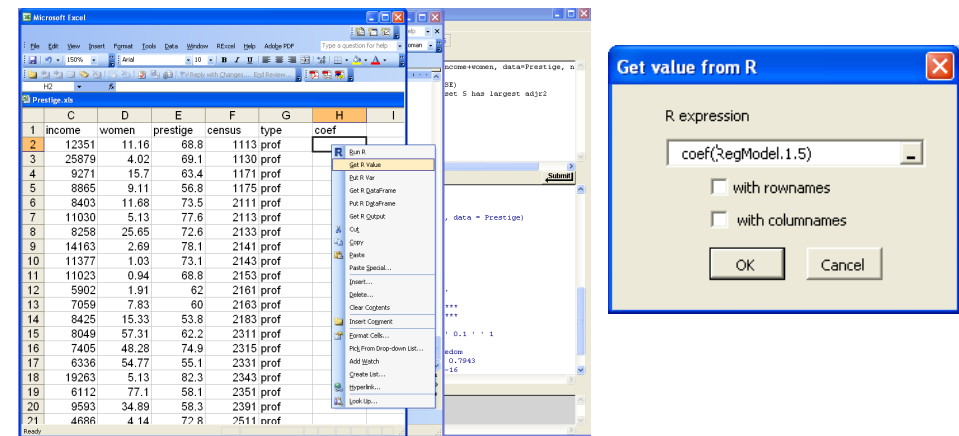
- The model specification box
 - Generates R script
 - Executes R script
 - Displays graph
 - Identifies model with the largest adjusted R^2

RExcel and RCommander



- displays the summary of the selected model

RExcel and RCommander



- We return the regression coefficients from the selected model back to the Excel spreadsheet.

RExcel and RCommander

	C	D	E	F	G	H	I
1	income	women	prestige	census	type	coef	
2	12351	11.16	68.8	1113	prof	-15.0731	
3	25879	4.02	69.1	1130	prof	0.000576	
4	9271	15.7	63.4	1171	prof	4.652593	
5	8865	9.11	56.8	1175	prof	0.0013	
6	8403	11.68	73.5	2111	prof		
7	11030	5.13	77.6	2113	prof		
8	8258	25.65	72.6	2133	prof		
9	14163	2.69	78.1	2141	prof		
10	11377	1.03	73.1	2143	prof		
11	11023	0.94	68.8	2153	prof		
12	5902	1.91	62	2161	prof		
13	7059	7.83	60	2163	prof		

- We return the regression coefficients from the selected model back to the Excel spreadsheet.

RExcel and RCommander

- Our addin to RCommander is currently available as an R package at <http://astro.ocis.temple.edu/~rmh/Rcmdr.HH>
- Some of our additions will be included in the next release of RCommander.
- We thank John Fox, the developer of RCommander, for help in designing our addins to his package.