



UseR! for Teaching

Sanford Weisberg
University of Minnesota
Minneapolis, MN USA

June 16, 2006



Audience: (Post)-graduates, both in statistics and particularly in other areas.

Non-stats student goals: Leave the class able to apply what they have learned to what they really care about.

Stats student goals: The material in the course, including the computing, is the end in itself.

Instructor's goals: Provide transferable knowledge, and keep computing from getting in the way (for non-stats students).

Three types of courses



- Teaching about R.
- Teaching *Analyzing Survey Data Using R*. This can imply teaching what the program can do under the general rubric of survey analysis.
- Using R in a course about sample surveys. This implies R is a tool that could be replaced by other tools.



- R provides a high-level language for research statisticians
- R is great for exploration of new ideas; packages.
- How to... courses, for example, graphics using R.
- Guru creation.

7



- Tailor the course to match what the program does. This often requires *compromise*.
- *Often, this is just what students want!*

“The University of Minnesota is not a technical or trade school.”

... Tom Burk, Forestry Prof.



- The program should enable, not hinder, learning methods. **Easy to say, hard to do.**
- Common metaphors for working with the computer are: browsers, iTunes, and possibly Excel. . . R is nothing like any of these and *therefore is not obvious to students.*
- Students get stuck on HOW rather than WHY; memorization (is it `header` or `col.names` or `colnames`?) and inconsistency are a hinderance.
- Irregular users *forget* — no visual cues: a blank screen is **intimidating.**
- Documentation is oriented toward the expert, not the novice (what is an S3 and why do I care?)



1999: Applied Regression Including Computing and Graphics

- Based on ARC and XLISPSTAT: Book and program are strongly linked: book and program inseparable: an intellectual success, but an overall failure.

2005: Applied Linear Regression, 3rd Ed

- Synthesis of last edition (1985), some graphics from 1999 book, and some new stuff
- Little mention of computing in the text.
- Web supplements for *ALR* using R, S-Plus, SAS, SPSS and JMP. (google applied linear regression).

Primer download statistics

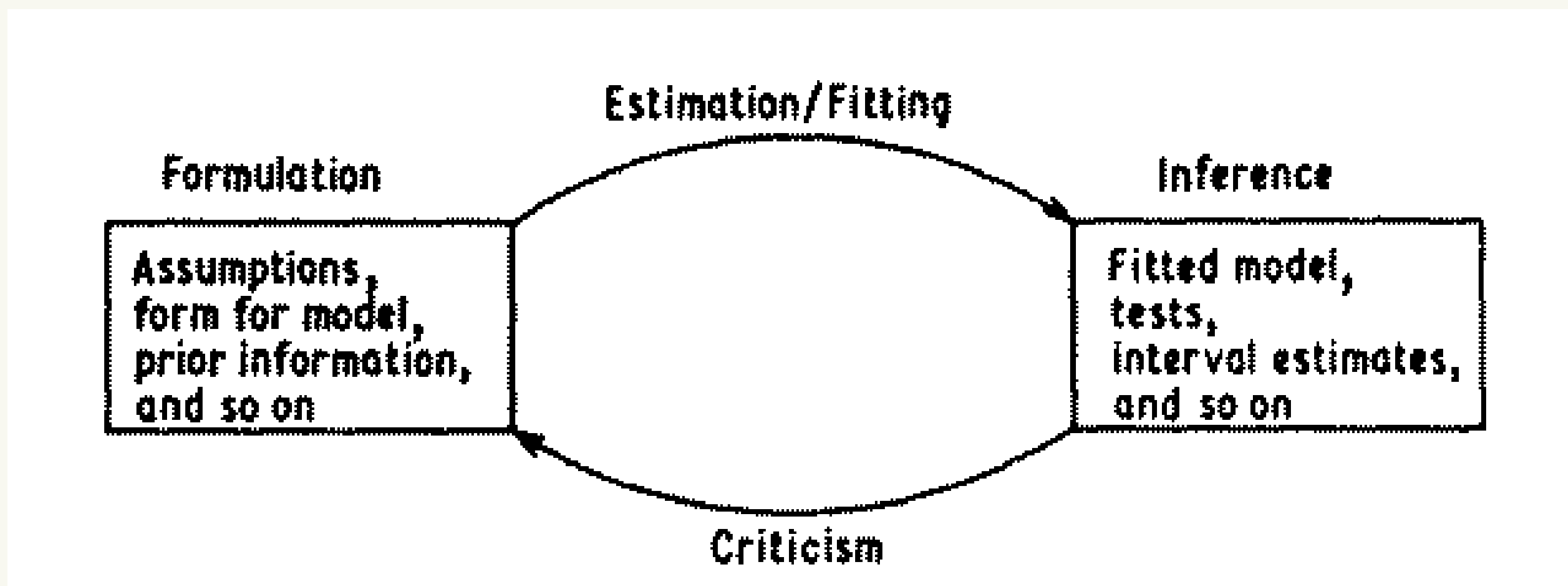


For January 1 – May 28, 2006, 11,000 web vists:

SPSS Primer	319	19%
SAS Primer	361	22%
JMP Primer	261	16%
R/SPlus Primer	725	44%

No program was adequate. R/S-Plus was closest with added package.

Does R encourage good data analysis?



If

```
> m1<-nls( y~th0+th1*(1-exp(-th2*x)),start=start)
```

How do you find `start`? How to chose the formula? What next?
Or, what before? How do you find out? No visual cues.



$$\theta_0 + \theta_1(1 - \exp(-\theta_2 D)), \text{ Deviance} = 3249.84$$

OLS NIL
▽

lowess NIL
▽

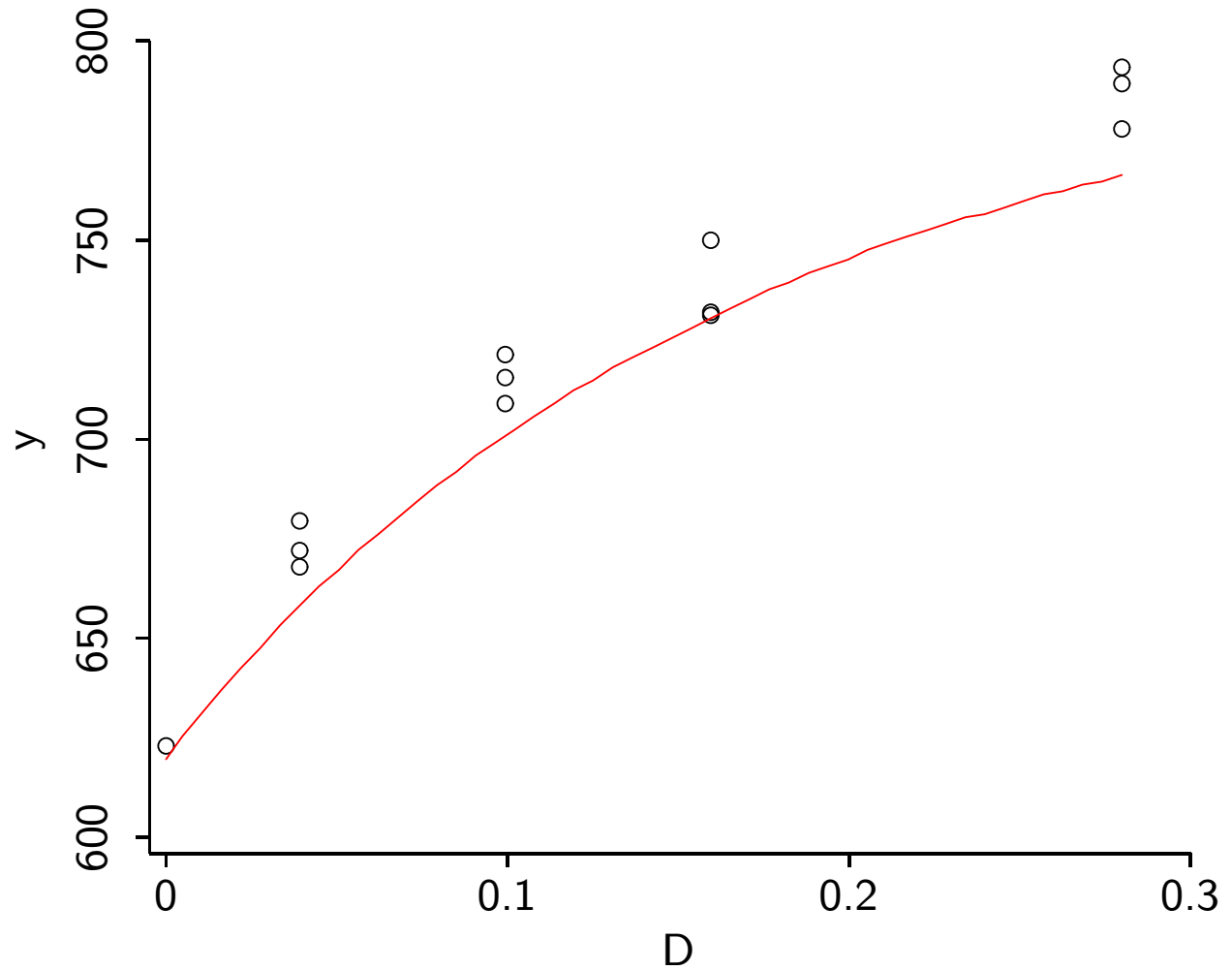
▽ Case deletions

TH0 620
▽

TH1 180
▽

TH2 6
▽

▽ More controls





- R works differently for different students, and R is unlikely to work for everyone.
- To help students:
 - Continued work on GUIs.
 - Improved, accessible documentation (Wiki).
 - Continued efforts to promote consistency that might be impossible with a commercial program but can be done in R.
 - Visual cues:



```
> library(alr3)
> m1 <- lm(LBM ~ Ht + Wt + RCC, data=ais)
> hints(m1)
```

Methods that understand lm objects:

conf.intervals	confidence intervals
inf.index	influence index plots
mmp	marginal model plots
pod	partial one-dimensional models
pure.error.anova	pure error analysis of variance
anova	analysis of variance
hatvalues	hat values
residual.plot	residual plotting methods
predict	predictions/fitted values
residuals	residuals of various types
inv.res.plot	inverse response plot
delta.method	estimate/se for nonlinear fns