

The use of Stata in Medical Statistics and Epidemiology: a long journey

Rino Bellocco, Sc.D.

Department of Statistics and Quantitative Methods
University of Milano-Bicocca

&

Department of Medical Epidemiology and Biostatistics
Karolinska Institutet

Nordic and Baltic Stata Users Group meeting, Stockholm,
September 1, 2017

Outline

A retrospective view

Stata Progress

Epidemiology and Biostatistics with Stata

My Journey with Stata

Outline

A retrospective view

Stata Progress

Epidemiology and Biostatistics with Stata

My Journey with Stata

Some History 1985-1995

- Version 1.0 - 1.5, January 1985 - February 1987 (around 48 commands, regress logit)

Basic Commands

Table 3: Stata 1.0 and Stata 1.1

append	dir	infile	plot	spool
beep	do	input	query	summarize
by	drop	label	regress	tabulate
capture	erase	list	rename	test
confirm	exit	macro	replace	type
convert	expand	merge	run	use
correlate	format	modify	save	
count	generate	more	set	
describe	help	outfile	sort	

Stata 1.0 had `int`, `long`, `float`, and `double` variables; it did not have `byte` or `str#`.

Some History 1985-1995 (cont)

- Version 2.0, June 1988, (Kaplan-Meier, Cox)
- *Statistics with Stata*, L. Hamilton, December 1989
- STB, May 1991
- Version 3.0, March 1993 (logistic, clogit, mlogit, ologit, epitab)
- Statalist, August 1994
- Version 4.0 (Windows 3.1), January 1995
- UK user's meeting, March 1995

Some History 1995-2005

- Version 5.0 , October 1996(xt, survey, st)
- Version 6.0, January 1999 (web, st)
- Version 7 (cluster, roc, smcl)
- StataCorp in College Station, December 2000
- GLM book, GLLAMM May-September 2001
- Macintosh, August 2001
- SJ, SE, November 2001 - February 2002
- An Introduction to Survival Analysis using Stata, May 2002
- Version 8 (Graphics, extended GUI)
- Measurement error software, September 2003
- Version 9, April 2005 (Mata, mixed)
- First Meeting in Sweden (Baltic and Nordic), September 2005

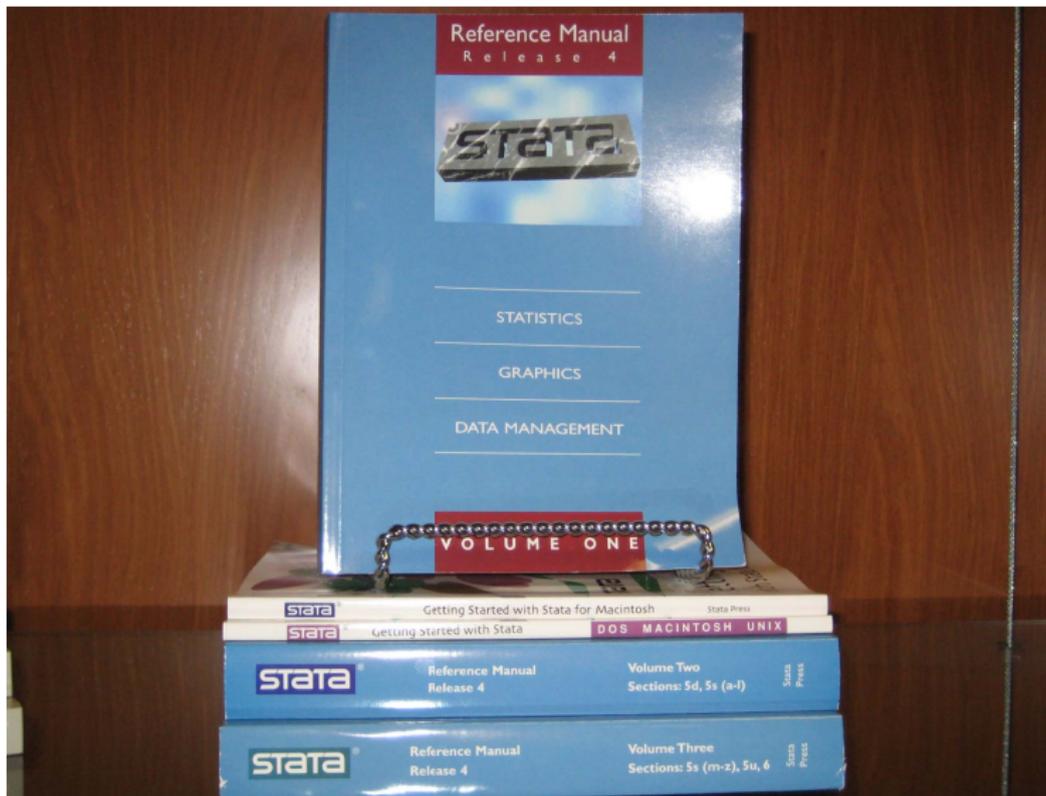
Some History 2006 -

- MP, October 2006
- Version 10, June 2007 (REM for logistic and poisson, exact logistic and poisson)
- Version 11, June 2009 (MI, GMM, CR, i., margin, pdf)
- Official Blog, December 2010
- Version 12, July 2011 (SEM, ROC, MI chained)
- Version 13, June 2013 (TE)
- Forum, good bye to Statalist@hsphsun2.harvard.edu
- Version 14, April 2015 (Bayes, ST treatment effects)
- Version 15, Jun 2017

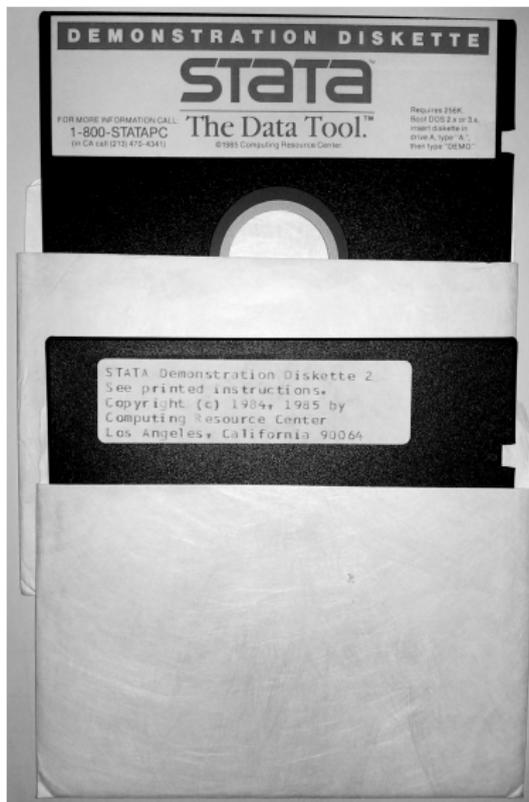
Manuals



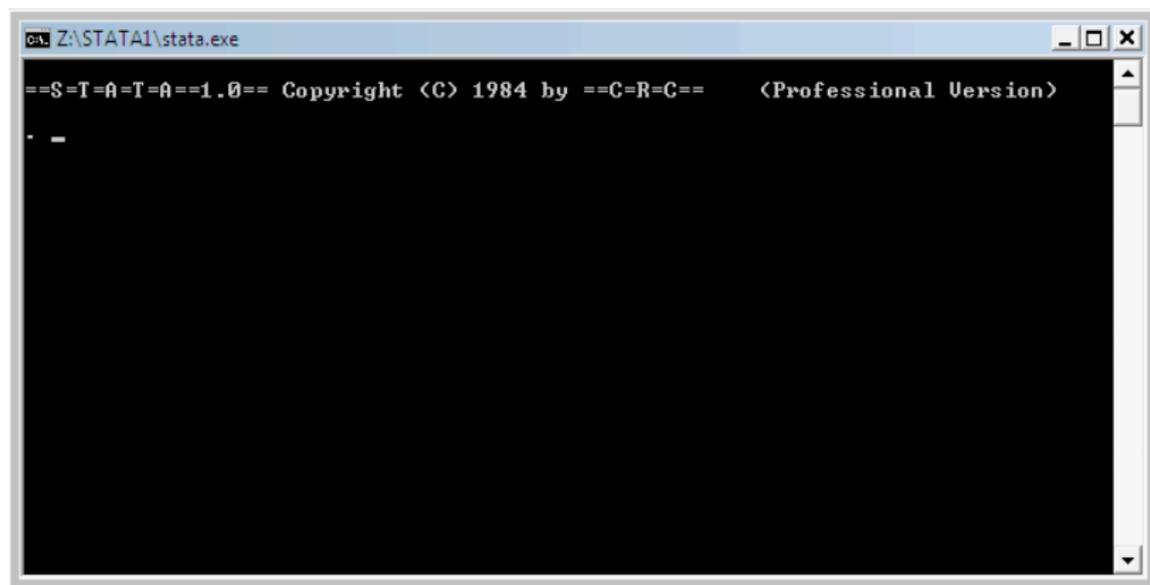
Manuals



Installation



The screen



The screenshot shows a Windows-style window titled "Z:\STATA1\stata.exe". The window contains a black background with white text. The text reads: "=="\$=T=A=T=A==1.0==" Copyright (C) 1984 by ==C=R=C== (Professional Version)". Below this, there is a small white cursor and a horizontal line. The window has standard Windows window controls (minimize, maximize, close) in the top right corner.

```
=="$=T=A=T=A==1.0==" Copyright (C) 1984 by ==C=R=C== (Professional Version)
```

Outline

A retrospective view

Stata Progress

Epidemiology and Biostatistics with Stata

My Journey with Stata

Version 1 - Version 2: Some Important Steps

- Survival Analysis, Kaplan-Meier estimate and Cox regression was implemented in its basic form in Version 2.
- The first Stata book was *Statistics with Stata*, from L. Hamilton
(Teaching of Statistics with Stata, or of Stata itself, has never been the first concern of Stata's developers, who see Stata as primarily software for research, but there has been always been strong and positive support for people who write texts linked to Stata (A quote from Nick Cox))
- The release of Stataques: the "student" statistical package, where the first version of the menu system was incorporated. Lawrence Hamilton suggested a "statistical calculator" in the program to take advantage of the "immediate" commands,

Version 3: much added

- logistic, ologit, oprobit, clogit, mlogit
- tobit, cnreg, rreg, qreg, weibull, ereg
- epitab
- weights reconsidered: pweights
- mvreg, sureg, heckman, nlreg, areg, canon
- nbreg
- constrained linear regression
- ml (first version)
- codebook

Version 3: EpiTab, a friend for epidemiologists

- Cohort studies
 - ir & iri
 - cs & csi
- Case control studies
 - cc & cci
 - tabodds
 - mhodds
- Matched Case control studies
 - cc & cci
 - tabodds
 - mhodds

Version 4 - Version 5

- Stata marries Windows
- Mac version released
- Stata.com
- glm
- xt commands introduced
- st package

Version 6

- web aware (use `http://`, updating, etc.)
- st rewritten

Version 7

- frailty
- cluster analysis
- roc
- Stata for MAC OS X
- Windowed UNIX
- *An Introduction to Survival Analysis using Stata*

Version 8

- graphics
- extended GUI, dialog boxes available for all commands
- gllamm
- simex

Version 9 - Version 10

- xtmixed
- mata
- improved do editor
- graphic editor
- exact methods
- power calculation

Version 11 - Version 12

- Multiple Imputation
- Competing Risk regression
- SEM
- MI chained
- ROC
- Margin
- pdf manuals

Version 13 - Version 15

- Treatment Effect
- Bayesian Analysis

Outline

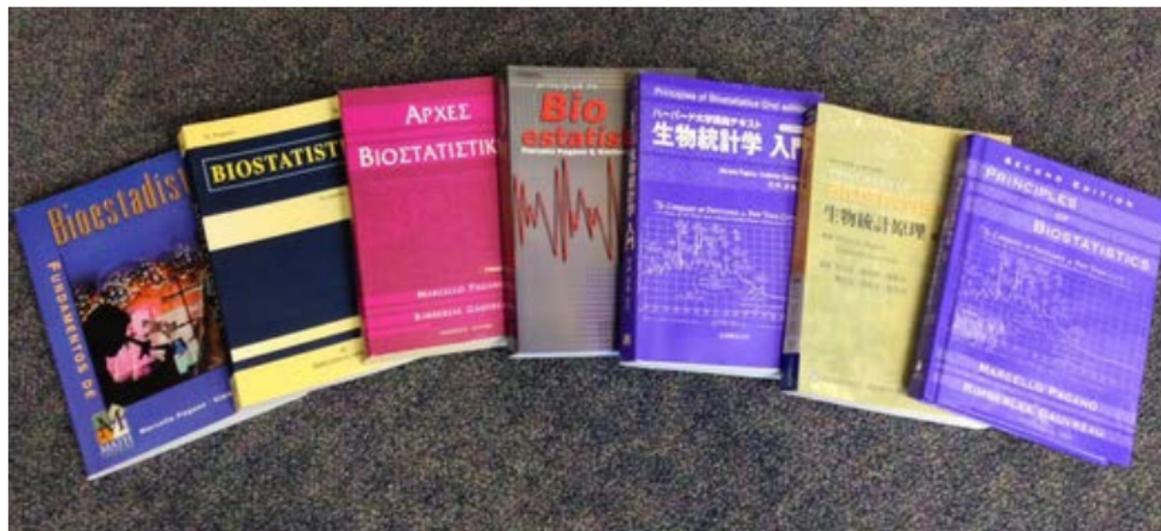
A retrospective view

Stata Progress

Epidemiology and Biostatistics with Stata

My Journey with Stata

The book(s) and the course



The book(s) and the course

THE WALL STREET JOURNAL.

WSJ.com

OPINION | October 2, 2012, 7:01 p.m. ET

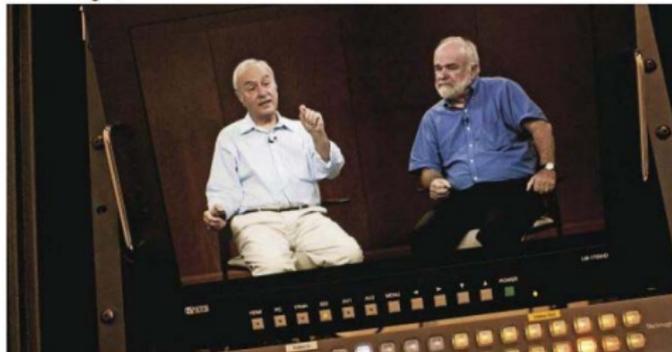
What Campuses Can Learn From Online Teaching

Harvard Online: A Worldwide Hit
By Kathleen Struck, Senior Editor, MedPage Today
Published: March 18, 2013

October 1, 2012

5 Ways That edX Could Change Education

By Marc Parry
Cambridge, Mass.



Kelvin Ma for the Chronicle

Teaching Material

- A short introduction to Stata for Biostatistics (Hill): *effmenu*
- An Introduction to Stata for Health Researchers (Juul)
- An handbook of Statistical Analysis using Stata (Everitt)
- Statistical Modeling for Biomedical Researchers (Dupont)
- Regression Methods in Biostatistics (Vittinghoff)
- Regression Models for Categorical Dependent Variables (Long)
- Generalized Linear Models and Extensions (Hardin)
- Multilevel and Longitudinal Modeling Using Stata, (Rabe-Hesketh)
- Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide (Twisk)
- An Introduction to Survival Analysis Using Stata (Cleves)
- Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model (Royston)
- Meta-Analysis in Stata: An Updated Collection from the Stata Journal (Sterne)

Where

- UCLA
- Harvard University, T.H. Chan School of Public Health
- John Hopkins, School of Public Health
- Karolinska Institutet
- London School of Tropical Medicine and Hygiene, Medical Statistics Unit
- University of Milano-Bicocca
- EW

Online resources

- <http://www.ats.ucla.edu/stat/stata/>

The screenshot shows a Mozilla Firefox browser window with the address bar containing <http://www.ats.ucla.edu/stat/stata/>. The page title is "Resources to help you learn and use Stata - Mozilla Firefox". The browser's toolbar includes navigation buttons (back, forward, home, stop, refresh) and a search bar with "Google" entered. The page content is organized into several sections:

Resources to help you learn and use Stata

Learning Stata

- [Stata Starter Kit](#)
- [Classes and Seminars](#)
- [Learning Modules](#)
- [Frequently Asked Questions](#)

Statistical Analysis

- [Data Analysis Examples](#)
- [Annotated Output](#)
- [Textbook Examples](#)
- [Paper Examples](#)
- [Web Books](#)
- [What statistical analysis should I use?](#)

Advanced Usage

- [Stata Programs for Research and Teaching](#)
- [Library](#)
- [Code Fragments](#)
- [Stata Tools for LaTeX](#)

Important Links

- [How can I get my own copy of Stata 10?](#)
- [Where to run Stata? How to get Stata?](#)
- [Installing, Customizing, Updating Stata](#)
- [Top 10 Questions about Stata](#)
- [The Stata Journal Archives](#)

Links by Topic

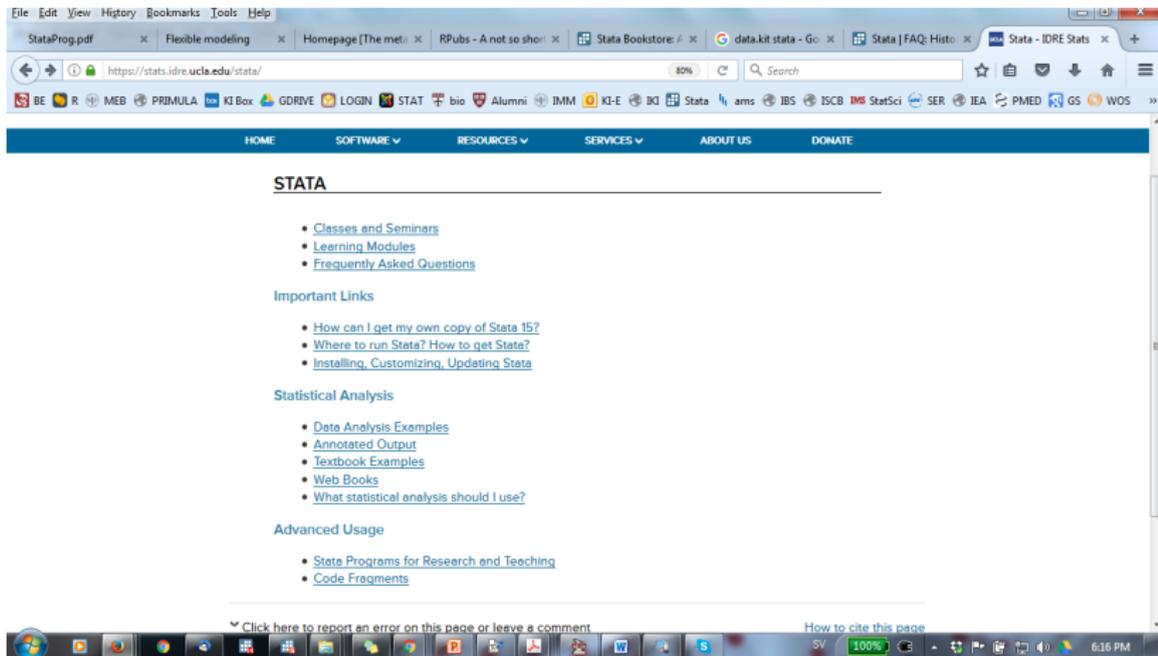
- [Stata 9 and 10](#)
- [Data Management](#)
- [Graphics](#)
- [ANOVA](#)
- [Regression](#)
- [Logistic \(and Categorical\) Regression](#)
- [Count Models](#)
- [Multilevel Modeling](#)
- [Survival Analysis](#)
- [Survey Data Analysis](#)

Lookup help for Stata command

Use [Google](#) to search our Stata pages for

Online resources

- <https://stats.idre.ucla.edu/stata/>



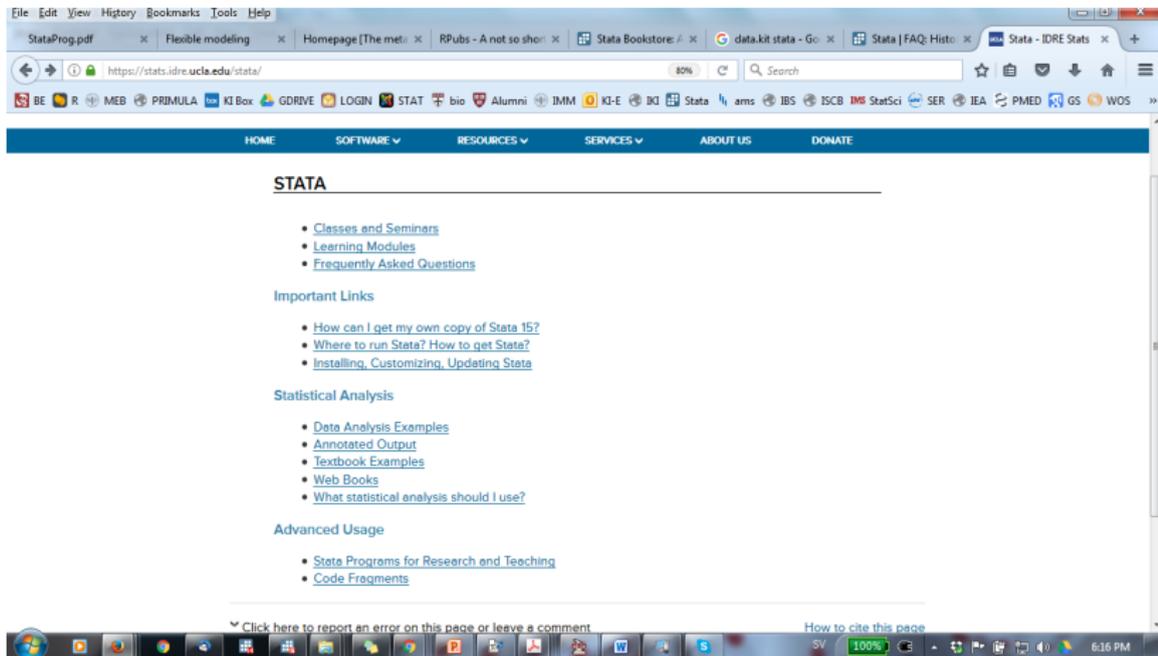
The screenshot shows a web browser window displaying the homepage of the Stata Institute's online resources. The browser's address bar shows the URL <https://stats.idre.ucla.edu/stata/>. The page features a blue navigation bar with links for HOME, SOFTWARE, RESOURCES, SERVICES, ABOUT US, and DONATE. Below the navigation bar, the word "STATA" is prominently displayed. The main content area is organized into several sections:

- STATA**
 - [Classes and Seminars](#)
 - [Learning Modules](#)
 - [Frequently Asked Questions](#)
- Important Links**
 - [How can I get my own copy of Stata 15?](#)
 - [Where to run Stata? How to get Stata?](#)
 - [Installing, Customizing, Updating Stata](#)
- Statistical Analysis**
 - [Data Analysis Examples](#)
 - [Annotated Output](#)
 - [Textbook Examples](#)
 - [Web Books](#)
 - [What statistical analysis should I use?](#)
- Advanced Usage**
 - [Stata Programs for Research and Teaching](#)
 - [Code Fragments](#)

At the bottom of the page, there are links to "Click here to report an error on this page or leave a comment" and "How to cite this page". The Windows taskbar is visible at the very bottom, showing the time as 6:16 PM.

Online resources

- <https://stats.idre.ucla.edu/stata/>



The screenshot shows a web browser window displaying the homepage of the Stata Institute's online resources. The browser's address bar shows the URL <https://stats.idre.ucla.edu/stata/>. The page features a blue navigation bar with links for HOME, SOFTWARE, RESOURCES, SERVICES, ABOUT US, and DONATE. Below the navigation bar, the word "STATA" is prominently displayed. The main content area is organized into several sections, each with a list of links:

- STATA**
 - [Classes and Seminars](#)
 - [Learning Modules](#)
 - [Frequently Asked Questions](#)
- Important Links**
 - [How can I get my own copy of Stata 15?](#)
 - [Where to run Stata? How to get Stata?](#)
 - [Installing, Customizing, Updating Stata](#)
- Statistical Analysis**
 - [Data Analysis Examples](#)
 - [Annotated Output](#)
 - [Textbook Examples](#)
 - [Web Books](#)
 - [What statistical analysis should I use?](#)
- Advanced Usage**
 - [Stata Programs for Research and Teaching](#)
 - [Code Fragments](#)

At the bottom of the page, there are two links: "Click here to report an error on this page or leave a comment" and "How to cite this page". The Windows taskbar is visible at the very bottom of the screenshot, showing the time as 6:16 PM.

Online resources

http://www.cpc.unc.edu/research/tools/data_analysis/statatutorial/

<http://www.michaelnormanmitchell.com/>

<http://data.princeton.edu/stata/>

<http://biostat.mc.vanderbilt.edu/wiki/Main/BiostatisticsTwoClassPage>

http://personalpages.manchester.ac.uk/staff/mark.lunt/stats_course.html

<https://www.iser.essex.ac.uk/resources/survival-analysis-with-stata>

<http://www.ssc.wisc.edu/sscc/pubs/stat.htm>

Outline

A retrospective view

Stata Progress

Epidemiology and Biostatistics with Stata

My Journey with Stata

Why Stata

- simple
- powerful
- cheap

When

- Dr. Pagano's meeting in 1992
- Version 2
- Impressed by simple clear command line window!
- Stata is introduced at KI in 1999, Niklas Berglind and I offering the first Stata course few years later with Nicola Orsini, supported by Peter H
- First SUG in Sweden
- Stata and the Summer School in Biostatistics and Epidemiology (from 1997)
- Stata 15

Harvard - 1993

- Most courses were already using Stata
- Bio210: Analysis of rates and proportions
- Bio220: Statistical Modeling
- Epi204: Analysis of Cohort and Case Control studies
- Bio216: Survival Analysis
- Bio221: Multivariate Categorical Data Analysis

Epidemiology and Biostatistics

- Epidemiological tables
- Survival analysis
- Linear, binary, and count regressions
- Survey methods
- Marginal means, contrasts, and interactions
- Multiple imputation
- Causal inference
- Multilevel mixed-effects models
- Bayesian analysis
- Programming and matrix programming

The importance of Stata Technical Bulletin

STATA
TECHNICAL
BULLETIN

September 1995

STB-27

A publication to promote communication among Stata users

Editor

Sean Beckett
Stata Technical Bulletin
8 Wakeman Road
South Salem, New York 10590
914-533-2278
914-533-2902 FAX
stb@stata.com EMAIL

Associate Editors

Francis X. Diebold, University of Pennsylvania
Joanne M. Garrett, University of North Carolina
Marcello Pagano, Harvard School of Public Health
James L. Powell, UC Berkeley and Princeton University
J. Patrick Royston, Royal Postgraduate Medical School

Subscriptions are available from Stata Corporation, email stata@stata.com, telephone 979-696-4600 or 800-STATAFPC, fax 979-696-4601. Current subscription prices are posted at www.stata.com/bookstore/stb.html.

Previous issues are available individually from StataCorp. See www.stata.com/bookstore/stbj.html for details.

Submissions to the STB, including submissions to the supporting files (programs, datasets, and help files), are on a nonexclusive, free-use basis. In particular, the author grants to StataCorp the nonexclusive right to copyright and distribute the material in accordance with the Copyright Statement below. The author also grants to StataCorp the right to freely use the ideas, including communication of the ideas to other parties, even if the material is never published in the STB. Submissions should be addressed to the Editor. Submission guidelines can be obtained from either the editor or StataCorp.

Copyright Statement. The Stata Technical Bulletin (STB) and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp. The contents of the supporting files (programs, datasets, and help files), may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the STB.

The insertions appearing in the STB may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the STB. Written permission must be obtained from Stata Corporation if you wish to make electronic copies of the insertions.

Users of any of the software, ideas, data, or other materials published in the STB or the supporting files understand that such use is made without warranty of any kind, either by the STB, the author, or Stata Corporation. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the STB is to promote free communication among Stata users.

The Stata Technical Bulletin (ISSN 1097-3879) is published six times per year by Stata Corporation. Stata is a registered trademark of Stata Corporation.

Contents of this issue

	page
an55. New Stata for Macintosh released	2
an56. Stata for Windows 95 and Stata for WindowsNT released	2
ip6. Repeat Stata command by variable(s)	3
snp6.2. Practical rules for bandwidth selection in univariate density estimation	5
ssa7. Analysis of follow-up studies	19
ssa8. Analysis of case-control and prevalence studies	26

The importance of Stata Technical Bulletin

General Categories:

an announcements

cc communications & letters

dm data management

dt data sets

gr graphics

in instruction

ip instruction on programming

os operating system, hardware, &
interprogram communication

qs questions and suggestions

tt teaching

zz not elsewhere classified

Statistical Categories:

sbe biostatistics & epidemiology

sed exploratory data analysis

sg general statistics

smv multivariate analysis

snp nonparametric methods

sqc quality control

sqv analysis of qualitative variables

srd robust methods & statistical diagnostics

ssa survival analysis

ssi simulation & random numbers

sss social science & psychometrics

sts time-series, econometrics

sxd experimental design

szz not elsewhere classified

- The importance of UK users (Michael and David to mention)
- Course in Finland in 2000: Statistical Practice in Epidemiology
- Our contribution from KI
- The interaction among users and with Stata
- The Stata Journal
- The Statalist (responsible for where I am)
- Where to go

Some nice commands in STB

- `ltable`
- The `lexis` procedure (precedes an older procedure by Hilbe)
- `tabodds`
- Updates (`lfit`, `lroc`)
- `lntrend`
- `xriml`
- `effmod`
- `pgmhaz`
- `concord`
- `epiconf`
- `univar`
- `oddsrcl`
- `stcascoh`
- `metaanalysis`

Thanks

- Marcello
- Bill
- Nicola

Thanks

