Introduction to Stata. Resources to help you learn and use Stata. UCLA: Statistical Consulting Group.
http://www.ats.ucla.edu/stat/stata/

The Stata Blog
http://blog.stata.com/

Cameron, A. C. and P. K. Trivedi (2009). Microeconometrics using Stata, Texas : Stata Press

> Microeconometrics Using Stata, by A. Colin Cameron and Pravin K. Trivedi, is an outstanding introduction to microeconometrics and how to do microeconometric research using Stata. Aimed at students and researchers, this book covers topics left out of microeconometrics textbooks and omitted from basic introductions to Stata. Cameron and Trivedi provide the most complete and up-to-date survey of microeconometric methods available in Stata. Early in the book, Cameron and Trivedi introduce simulation methods and then use them to illustrate features of the estimators and tests described in the rest of the book. While simulation methods are important tools for econometricians, they are not covered in standard textbooks. By introducing simulation methods, the authors arm students and researchers with techniques they can use in future work. Cameron and Trivedi address each topic with an in-depth Stata example, and they reference their 2005 textbook, Microeconometrics: Methods and Applications, where appropriate. The authors also show how to use Stata's programming features to implement methods for which Stata does not have a specific command. Although the book is not specifically about Stata programming, it does show how to solve many programming problems. These techniques are essential in applied microeconometrics because there will always be new, specialized methods beyond what has already been incorporated into a software package. Cameron and Trivedi's choice of topics perfectly reflects the current practice of modern microeconometrics. After introducing the reader to Stata, the authors introduce linear regression, simulation, and generalized least-squares methods. The section on cross-sectional techniques is thorough, with up-to-date treatments of instrumental-variables methods for linear models and of quantile-regression methods. The next section of the book covers estimators for the parameters of linear panel-data models. The authors? choice of topics is unique: after addressing the standard random-effects and fixed-effects methods, the authors also describe mixed linear models'a method used in many areas outside of econometrics. Cameron and Trivedi not only address methods for nonlinear regression models but also show how to code new nonlinear estimators in Stata. In addition to detailing nonlinear methods, which are omitted from most econometrics textbooks, this section shows researchers and students how to easily implement new nonlinear estimators. The authors next describe inference using analytical and bootstrap approximations to the distribution of test statistics. This section highlights Stata's power to easily obtain bootstrap approximations, and it also introduces the basic elements of statistical inference. Cameron and Trivedi then include an extensive section about methods for different nonlinear models. They begin by detailing methods for binary dependent variables. This section is followed by sections about multinomial models, tobit and selection models, count-data models, and nonlinear panel-data models. Two appendices about Stata programming complete the book.

Cleves, M., et al. (2008). An introduction to survival analysis using Stata, Texas : Stata Press

> This book is the ideal tutorial for professional data analysts who want to learn survival analysis for the first time or who are well versed in survival analysis but not as dexterous in using Stata to analyze survival data. This text also serves as a valuable reference to those who already have experience using Stata's survival analysis routines. The second edition has been updated for Stata 10, containing a new chapter on power and sample-size calculations for survival studies and sections that describe how to fit regression models (stcox and streg) in the presence of complex survey data. Other enhancements include discussions about nonparametric estimation of mean/median survival, survival graphs with embedded at-risk tables, better hazard graphs through the use of boundary kernels, and concordance measures for assessing the predictive accuracy of the Cox model, as well as an expanded discussion of model building strategies including the use of fractional polynomials. Survival analysis is a field of its own requiring specialized data management and analysis procedures. Toward this end, Stata provides the st family of commands for

organizing and summarizing survival data. The authors of this text are also the authors of Stata's st commands. This book provides statistical theory, step-by-step procedures for analyzing survival data, an in-depth usage guide for Stata's most widely used st commands, and a collection of tips for using Stata to analyze survival data and present the results. This book develops from first principles the statistical concepts unique to survival data and assumes only a knowledge of basic probability and statistics and a working knowledge of Stata. The first three chapters of the text cover basic theoretical concepts: hazard functions, cumulative hazard functions, and their interpretations; survivor functions; hazard models; and a comparison of nonparametric, semiparametric, and parametric methodologies. Chapter 4 deals with censoring and truncation. The next three chapters cover the formatting, manipulation, stsetting, and error checking involved in preparing survival data for analysis using Stata's st analysis commands. Chapter 8 covers nonparametric methods, including the Kaplan-Meier and Nelson-Aalen estimators, and the various nonparametric tests for the equality of survival experience. Chapters 9-11 discuss Cox regression and include various examples of fitting a Cox model, obtaining predictions, interpreting results, building models, and model diagnostics. The next four chapters cover parametric models, which are fit using Stata's streg command. These chapters include detailed derivations of all six parametric models currently supported in Stata and methods for determining which model is appropriate, as well as information on obtaining predictions, stratification, and advanced topics such as frailty models. The final chapter is devoted to power and sample-size calculations for survival studies.

Hesketh, S. and A. Skrondal (2008). Multilevel and longitudinal modeling using Stata, Texas : Stata Press

Multilevel and Longitudinal Modeling Using Stata, by Sophia Rabe-Hesketh and Anders Skrondal, looks specifically at Stata's treatment of generalized linear mixed models, also known as multilevel or hierarchical models. These models are "mixed" because they allow fixed and random effects, and they are "generalized" because they are appropriate for continuous Gaussian responses as well as binary, count, and other types of limited dependent variables. Beginning with the comparatively simple random-intercept linear model without covariates, Rabe-Hesketh and Skrondal develop the mixed model from principles, thereby familiarizing the reader with terminology, summarizing and relating the widely used estimating strategies, and providing historical perspective. Once the authors have established the mixed-model foundation, they smoothly generalize to random-intercept models with covariates and then to random-coefficient models. The middle chapters of the book apply the concepts for Gaussian models to models for binary responses (e.g., logit and probit), ordinal responses (e.g., ordered logit and ordered probit), and count responses (e.g., Poisson). The authors then consider models with multiple levels of random variation and models with crossed (nonnested) random effects. In its examples and end-of-chapter exercises, the book contains real datasets and data from the medical, social, and behavioral sciences literature. The book has several applications of generalized mixed models performed in Stata. Rabe-Hesketh and Skrondal developed gllamm, a Stata program that can fit many latent-variable models, of which the generalized linear mixed model is a special case. With the release of version 9, Stata introduced the xtmixed command for fitting linear (Gaussian) mixed models. Stata users can use gllamm and xtmixed in conjunction with the rest of the xt suite of commands to perform comparative mixed-model analyses for various response families. The type of models fit by these commands sometimes overlap; when this happens, the authors highlight the differences in syntax, data organization, and output for the two (or more) commands that can be used to fit the same model. The authors also point out the relative strengths and weaknesses of each command when used to fit the same model, based on considerations such as computational speed, accuracy, and available predictions and postestimation statistics. The book delineates the relationship between gllamm and xtmixed, clearly showing how they complement one another.

Blossfeld, H. P., et al. (2007). Event history analysis with Stata, New York : Lawrence Erlbzum Associates

This book presents survival analysis from a social science perspective. Introducing the mathematics and statistics of survival analysis, along with substantive discussions of social science data?specific issues, the authors give examples throughout using Stata (version 9) and data from the German Life History Study.

The text covers both basic and advanced topics, from an introduction to life tables to fitting parametric models with unobserved heterogeneity. The authors aptly illustrate the entire research path required in applying event history analysis, from the initial problems of recording event-oriented data, to data organization, to software applications, to interpreting results. Chapters 1 and 2 introduce event history data, discussed substantively, and the data structures used to contain them. Chapter 3 introduces nonparametric descriptive methods including life tables, product-limit estimation of the survivor function, and comparison of survivor functions. Chapters 4-8 focus on estimation using parametric survival functions. This section discusses not the usual exponential, Weibull, etc., models but rather issues such as period-specific effects, qualitative and quantitative covariates, time-dependent covariates, and multiepisode data. Chapter 9 discusses the Cox proportional hazards model, whereas chapter 10 covers problems with parametric model specification, including unobserved heterogeneity. The book has a parametric model focus, which for some readers will be a strength and for others, a weakness. For the latter group, the weakness is minimal because the coverage of the Cox model is adequate given the foregoing discussion. Event History Analysis with Stata is aimed at the professional social scientist but could also serve as a graduate-level text. A web site providing supporting materials for the book, including the dataset files and do-files, is available at http://web.uni-bamberg.de/sowi/soziologie-i/eha/stata.

Gould, W., et al. (2006). Maximum likehood estimation with Stata, Texas : Stata Press

Maximum Likelihood Estimation with Stata is written for researchers in all disciplines who need to fit models using maximum likelihood estimation. This edition offers a wealth of material about the ml command, updated to include new features introduced in Stata 9.

Scott, Long J. and J. Freese (2006). Regression models for categorical dependent variables using Stata, Texas : Stata press

This book shows how to use Stata to fit and interpret regression models for categorical data. Nearly 50% longer than the previous edition, the second edition covers new topics for fitting and interpreting models included in Stata 9, such as multinomial probit models, the stereotype logistic model, and zero-truncated count models. Many of the interpretation techniques have been updated to include interval and point estimates. Although regression models for categorical dependent variables are common, few texts explain how to interpret such models; Regression Models for Categorical Dependent Variables Using Stata, Second Edition fills this void. To accompany the book, Long and Freese provide a suite of commands for hypothesis testing and model diagnostics. The second edition begins with an excellent introduction to Stata and follows with general treatments of estimation, testing, fit, and interpretation in this class of models. Long and Freese detail binary, ordinal, nominal, and count outcomes in separate chapters. The final chapter explains how to fit and interpret models with special characteristics, such as interaction, nonlinear terms, and ordinal and nominal independent variables. One appendix explains the syntax of the author-written commands, and a second appendix details the book's datasets. Long and Freese use many concrete examples in their second edition. All the examples, datasets, and author-written commands are available on the authors' website, so readers can easily replicate the examples when using Stata. This book is ideal for students or applied researchers who want to learn how to fit and interpret models for categorical data (4e de couverture).

La statistique avec R

The R Project for Statistical Computing. Documentation.
https://www.r-project.org/

The Comprehensive R Archive Network. Documentation.
https://cran.r-project.org/

W.N. Venebles & D. M. Smith, An Introduction to R
http://cran.r-project.org/doc/manuals/R-intro.pdf
http://sites.google.com/site/r4statistics/books/free-version

Husson, F., et al. (2009). Analyse de données avec R, Rennes : Presses Universitaires de Rennes

> Avec ce manuel, le lecteur dispose d'un équipement complet (bases théoriques, exemples, logiciels) pour analyser des données multidimensionnelles. Pour chaque méthode, un exemple détaillé concrétise les éléments théoriques et chaque résultat est accompagné de la commande R qui permet de l'obtenir (logiciel FactoMineR).

Jérôme Pagès (2014). Multiple Factor Analysis by Example Using R