

The Top Resources for Learning 13 Statistical Topics

By Karen Grace-Martin
Founder & President



About the Author



Karen Grace-Martin is the founder and president of The *Analysis Factor*. She is a professional statistical consultant with Masters degrees in both applied statistics and social psychology. Her career started in psychology research, where her frustration in applying statistics to her data led her to take more and more statistics classes. She soon realized that her favorite part of research was data analysis, leading to a career change. Her background in experimental research and working with real data has been invaluable in understanding the challenges that researchers face in using statistics and has spurred her passion for deciphering statistics for academic researchers.

Karen was a professional statistical consultant at Cornell University for seven years before founding The Analysis Factor. Karen has worked with clients from undergraduate honors students on their first research project to tenured Ivy League professors, as well as non-profits and businesses. Her ability to understand what researchers need and to explain technical information at the level of the researcher's understanding has been one of her strongest assets as a consultant. She treats all clients with respect, and derives genuine satisfaction from the relief she hears in their voices when they realize that someone can help them.

Before consulting, Karen taught statistics courses for economics, psychology, and sociology majors at the University of California, Santa Barbara and Santa Barbara City College. Karen has also developed and presented many statistics workshops, most recently on missing data, logistic regression, and interpreting regression parameters.

Karen has co-written an introductory statistics textbook with sociologist Stephen Sweet: *Data Analysis with SPSS*. It focuses on statistical concepts and data analysis practices, without the endless calculations that often obscure them.

Introduction

I have devoted my entire professional life to excellence in research and data analysis in academic settings. My passion is in helping academic researchers learn, apply, and practice statistics without the frustration that so often comes with it.

One role I have that is of great help to researchers is recommending good resources. Let's face it—there is a whole lot of information out there about statistics, and much of it is incomprehensible. But there are gems in the rough. This paper is a collection of the best resources I can recommend to researchers who need to learn a new topic in statistics.

I assume you do have a good basic statistical base—these won't be helpful for newbies. You need to have a good conceptual understanding of hypothesis testing, general statistical concepts, and be relatively comfortable performing a multiple regression or factorial ANOVA.

Most academic researchers I work with are at this level, but now need to go beyond it without taking more statistics classes. These resources will help you learn and apply statistical methods that are at one level beyond the general linear model.

I have chosen the resources that I believe are most helpful for learning that topic. These lists are not exhaustive—you can find many more resources at <http://analysisfactor.com/resources/resources.html> and at my blog: <http://www.analysisfactor.com/statchat>. Here I've picked the best resources for each one. These are the ones I would start with if I were learning or needing to refresh my understanding of a topic.

That said, they are chosen for their readability as well. I don't believe that being adept at regression requires you understand linear algebra or means you find it helpful to see equations or statistical theory derived. When possible, I have chosen non-technical sources. That doesn't mean some won't be a stretch. Read it twice, or even three times. Try out the method, then read it again.

Some topics have more resources listed than others. Some topics need only one, some topics have only one. But the list in each topic should be enough to get you started doing that method well.

I am always interested in hearing about other good resources for learning and applying statistics, so if you know of any, send me an email please. I consider this ebook a work in progress!

Table of Contents

Complex Surveys	5
Factor Analysis	5
Logistic Regression	5
Missing Data	6
Mixed ANOVA Models	7
Multilevel Regression Models	7
Nonparametric Statistics	8
Path Analysis	8
Poisson & Negative Binomial Regression	8
Principal Components Analysis	9
Sample Size Calculations	9
Structural Equation Modeling	9
Survival and Event History Analysis	10
Tobit Regression	10

Complex Surveys

[Pitfalls of Using Standard Statistical Software Packages for Sample Survey Data.](#)

by Donna J. Brogan

[Analysing Complex Survey Data: Clustering, Stratification and Weights](#)

by Patrick Sturgis

[Summary of Survey Analysis Software](#)

By Harvard University Survey Research Methods Section

<http://www.hcp.med.harvard.edu/statistics/survey-soft/>

Factor Analysis

[An Explanation of Factor Analysis](#)

by Richard Darlington, Cornell University

<http://comp9.psych.cornell.edu/Darlington/factor.htm>

[A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling](#)

by Larry Hatcher

This is THE book to read if you're doing Factor Analysis. The first chapter alone on Principal components is amazing. Very well written and nontechnical. Even if you don't use SAS, this book explains Factor Analysis VERY well. Just ignore the SAS code.

Logistic Regression

DeMaris, Alfred. (1995) **A Tutorial in Logistic Regression.** *Journal of Marriage and the Family*, 57, 956-968.

Morgan, Philip S., Teachman, Jay D. (1988) **Logistic Regression: Description, Examples, and Comparisons.** *Journal of Marriage and the Family*, 50, 929-936.

[Logistic Regression Using the SAS System : Theory and Application](#)

by Paul Allison

You've probably noticed by now that I really like Paul Allison's books. He is one of my favorite applied statistics authors. This one is great as well.

[Applied Logistic Regression Analysis](#)

by Scott Menard

A great introduction to logistic regression, and very reasonably priced. I love all of the books in this series that I've read. Contains enough statistical theory to be helpful, but not lots of technical jargon or deriving equations. Readers should be quite familiar with linear regression.

Missing Data

Allison, Paul (2000). [Multiple Imputation for Missing Data: A Cautionary Tale](#). *Sociological Methods and Research*, 28, 301-309.

Schafer, J.L. & Graham, J.W. (2002). **Missing Data: Our View of the State of the Art**, *Psychological Methods*, 7, 147-177.

[Missing Data](#)

by Paul Allison

Very reader-friendly. One of "the little green Sage books." This is an excellent overview, covers much of what a data analyst needs to know, and very accessible. This is the book to start with. And very reasonably priced.

<http://multiple-imputation.com>

Mixed ANOVA Models

[SAS for Mixed Models](#), Second Edition

by Ramon Littell, George Miliken, Walter Stroup, Russell Wolfinger, & Oliver Schabenberger

This is a pretty technical book, and is not for the statistically feeble. That said, if you are doing Mixed Modeling in SAS, it's a must-have. Back in the Cornell Statistical Consulting office, we actually wore ours out.

[Mixed Up Mixed Models: Things That Look Like They Should Work But Don't And Things That Look Like They Shouldn't Work But Do](#)

By Robert Hamer & P.M. Simpson

<http://www2.sas.com/proceedings/sugi25/25/aa/25p020.pdf>

Multilevel Regression Models

Arnold, Carolyn L. (1992). **An Introduction to Hierarchical Linear Models**. *Measurement and Evaluation in Counseling and Development*, 25, 58-90.

Plewis, Ian. (1998) [Multilevel Models](#). *Social Research Update*, 23.
<http://sru.soc.surrey.ac.uk/SRU23.html>

Singer, Judith D. (1998) [Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models](#). *Journal of Educational and Behavioral Statistics*, 24, 323-355.
<http://gseweb.harvard.edu/~faculty/singer/Papers/Using%20Proc%20Mixed.pdf>

This article uses SAS, but the ATS group at UCLA has on their web site papers that go through the examples in [HLM](#), [MLwiN](#), [Stata](#), [SPSS](#), and [SPLUS](#).
<http://www.ats.ucla.edu/stat/paperexamples/singer/default.htm>

SAS for Mixed Models, Second Edition

by Ramon Littell, George Miliken, Walter Stroup, Russell Wolfinger, & Oliver Schabenberger

This is a pretty technical book, and is not for the statistically feeble. That said, if you are doing Mixed Modeling in SAS, it's a must-have. Back in the Cornell Statistical Consulting office, we actually wore ours out.

Nonparametric Statistics

StatXact 3 for Windows: Statistical Software for Exact Nonparametric Inference, Quick Reference Guide

There are quite a few books on nonparametric statistics, and they're not necessarily difficult, but this book is the best. It is actually a software manual for StatXact, which does exact p-value calculations on a huge number of non-parametric statistics. It has, however, amazingly clear explanations on just about every nonparametric test you'd ever want to do.

Path Analysis

Baron, Reuben M. & Kenny, David A. **The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations.** *Journal of Personality and Social Psychology*, 51, 1173-1182

Poisson & Negative Binomial Regression

Gardener, W., Mulvey, E.P., Shaw, E.C. (1995) **Regression Analyses of Counts and Rates: Poisson, Overdispersed Poisson, and Negative Binomial Models.** *Psychological Bulletin*.

Long, J.S. (1997) **Regression Models for Categorized and Limited Dependent Variables.** Thousand Oaks, CA: Sage Publications.

Principal Components Analysis

Chapter 1 of: [A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling](#)

by Larry Hatcher

This is by far the best explanation of Principal Components Analysis (with step-by-step instructions) I've seen, even if you don't use SAS.

[StatNews 49: Principal Components: Not Just Another Factor Analysis](#)

By Yun Wang

<http://cscu.cornell.edu/news/statnews/stnews49.pdf>

Sample Size Calculations

Hoenig, John M. and Heisey, Dennis M. (2001), ``**The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis**,'' *The American Statistician*, **55**, 19-24.

Lenth, R. V. (2001), [Some Practical Guidelines for Effective Sample Size Determination](#), *The American Statistician*, **55**, 187-193.

Lenth, R. V. (2002). [Two Sample Size Practices that I don't Recommend](#), Presented at the Joint Statistical Meetings.

Structural Equation Modeling

[Ed Rigdon's SEM FAQ](#)

[Structural Equations with Latent Variables](#)

By Kenneth Bollen

This is dry, and not light reading, but it is the definitive text.

Survival and Event History Analysis

Singer, J.D. & Willet, J.B. (1993). *It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events*. Journal of Educational Statistics, 18, 155-195.

[Event History Analysis: Regression for Longitudinal Event Data](#)

By Paul Allison

[Survival Analysis Using SAS: A Practical Guide](#)

by Paul Allison

This book alone will teach you just about everything you need to know about Survival Analysis. It's worthwhile, even if you don't use SAS.

Tobit Regression

[Regression Models: Censored, Sample Selected, or Truncated Data](#)

by Richard Breen

This is a pretty specific topic, but it's a great book when you need it. I love all of the books in this series that I've read. Readers should be quite familiar with linear regression.