APPENDIX A

FURTHER READING

A.1 Introduction

Given the ubiquity of statistical analysis, there are an enormous number of books on how to analyze data, many of them aimed at particular specialties ranging from wildlife biology to finance – and varying in mathematical level from the extremely elementary to the very advanced. This chapter represents our attempt to suggest useful books for particular areas, and also general books that we have found particularly useful.

Two of the books cited below (*Silverman* [1986] and *Efron and Tibshirani* [1993]) are part of a series of short statistical monographs written for potential users of statistical methods rather just as expositions of the theory. Given a topic, it is worthwhile seeing if there is a book in this series that covers it. The journal literature in statistics, like that in any field, is mostly entirely papers written for the other dozen experts on the topic, and comprehensible to another few hundred. An occasional exception is the review articles in the journal *Statistical Science*.

A.2 Probability and Statistics: General

Bevington and Robinson [2003] is a useful manual on the treatment of measurement errors in physical data; it has long served (in earlier editions) as the standard book for statistics in physics labs. The emphasis is on Gaussian statistics, but it also includes error propagation, least squares and maximum likelihood estimation, and curve fitting.

Hogg and Tanis [2000] is a good example of the class of relatively-sophisticated introductory statistics textbooks, covering all the basic areas. We prefer *Rice* [1995] (now out of print; *Rice* [2006] is not, but is stunningly expensive); this takes a somewhat different viewpoint, perhaps more up-to-date. *Dekking et al.* [2005] is more elementary but covers a broader range

of topics. *Sivia* [1996] is a good introduction to statistical analysis with a different flavor than others, namely the Bayesian approach, which is becoming more and more popular. *Evans et al.* [2000] is a handy reference on univariate probability density functions, with plots, summary information, and instructions on how to generate random variates.

More of a reference work, though still containing exercises, and laid out fairly didactically, are the books by *Stuart and Ord* [1991, 1994], which are called *Kendall's Advanced Theory of Statistics* because they are descendants of *The Advanced Theory of Statistics*, by M. Kendall and Stuart; the original was in three volumes, the third of which has not yet been revised. The treatment is not what modern mathematicians would call advanced, but it is very thorough: you can expect to find a discussion of almost any topic. The sections on inference are quite readable, though probably better read after you have had some experience than as an introduction.

Wall and Jenkins [2012] introduces a variety of statistical methods for astronomical data – many of which, because astronomy and geophysics are both observational sciences, are often useful for geophysicists. The focus of the second edition has moved much more towards Bayesian methods; though the book is quite well-written, it is aimed more at the practitioner who wants to learn methods than at students learning the subject for the first time.

A.3 Special Topics

Efron and Tibshirani [1993] is what it says, an introduction to the bootstrap and jackknife methods, at a level that is moderately advanced but still approachable. *Davison and Hinkley* [1997] is a more advanced treatment, which presupposes that you already know what is in the Efron and Tibshirani book.

Fisher [1993] covers statistical methods for data that are defined on a circle (that is, directions in two dimensions), while *Fisher et al.* [1987] and *Tauxe* [1998] cover those for data on a sphere: both situations of interest in geophysics.

Seber and Lee [2003] is a thorough treatment of regression analysis, that is fitting known functions to random variables. The emphasis is on the statistical aspects, rather than the numerical ones. *Lancaster and Salkauskas* [1986] describes methods of curve fitting, emphasizing the numerical aspects, and also the exact fitting of curves to data. *Silverman* [1986] pro-

vides a very readable discussion of how to summarize data by a density function. If you look through these three books, you should be able to find something appropriate for fitting a function to your dataset.

Cox and Lewis [1966] remains a good introduction to the analysis of point processes.

A.4 Computational Statistics

This is, of course, the most difficult area to give up-to-date references for, given the proliferation of software packages and free software on the Web. Many of the commercial statistics packages are designed to allow people to make computations without a lot of understanding, which is hardly what we would recommend. One option is *MATLAB*, which contains procedures for most of the activities covered in this course. But we urge that for data analysis you should seriously consider `R`, which is open-source software[1] that has become the most-used package for statisticians. Most new results in statistical computation are presented as scripts in this language, so it includes a large number of procedures not otherwise conveniently available;

However, you may eventually need to use a programming language in which case it is good to have subroutines available. One place to look is *Press et al.* [1992], which (unlike most numerical analysis books), covers data analysis methods, including a number of statistical computations – though, reasonably enough, they focus on the computational aspects rather than the statistical concepts. You should also look at the collection of routines hosted by Statlib[2] taken from algorithms published by the journal *Applied Statistics*. *Gershenfeld* [1999] has some probability and statistics, as well as a good treatment of optimization, while covering a whole range of useful mathematical methods, from the traditional and analytical to the very modern and heavily computational. Since geophysics is all about mathematical modeling, this is a book you may wish to have as a reference.

[1] Available at `http://www.r-project.org`.

[2] `http://lib.stat.cmu.edu/apstat/)`

A.5 Data Presentation

The graphical display of data is important at both ends of a statistical analysis. You should start by looking at your data, and will end by trying to show others what your analysis means, and both of these aims are best met by making some kind of plot. *Cleveland* [1993] describes graphical methods for looking at data, especially multivariate data, while *Cleveland* [1985] discusses how plots can be more effective in communicating their contents – or for that matter less effective, as he shows lots of bad examples. The importance of data display for geophysics is nicely illustrated by a graph on page 145 of *Cleveland* [1985], which shows that, out of 57 journals surveyed, *J. Geophys. Res.* had the highest proportion of page space devoted to graphs: over 30%.

Tufte [2001, 1991, 1997] are three beautifully produced books on the overall theme of how to present data (and other things) in visually effective ways. The 2001 book (a second edition) is most relevant to displaying data, but all three books are great sources of ideas and cautionary points.[3] You should read them all – and then read his diatribe on the awfulness of PowerPoint.

A.6 History

Gigerenzer et al. [1989] and *Salsburg* [2001] describe the (rather convoluted) history of this pair of subjects, and reflect on what it is all about in the process. *Gigerenzer et al.* [1989] covers the whole period from the 17th century to the present; the chapter on statistical inference is worth reading even if you have no interest in the history. Salsberg's book is restricted to the last century, with many vignettes of statisticians, and an interesting viewpoint (he works in medical statistics).

[3] There is also a fourth, which we do not think is so useful.

Bibliography

Agnew, D. C., Comment on "Changes of reporting rates in the Southern California earthquake catalog, introduced by a new definition of ML" by Thessa Tormann, Stefan Wiemer, and Egill Hauksson, *Bull. Seismol. Soc. Am.*, *100*, 3320–3324, doi:10.1785/0120100027, 2010.

Agnew, D. C., and L. M. Jones, Prediction probabilities from foreshocks, *J. Geophys. Res.*, *96*, 11,959–11,971, 1991.

Bakun, W., and T. V. McEvilly, Recurrence models and Parkfield, California, earthquakes, *J. Geophys. Res.*, *89*, 3051–3058, 1984.

Bevington, P. R., and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, Dubuque, Iowa, 2003.

Bird, P., An updated digital model of plate boundaries, *Geochem. Geophys. Geosys.*, *4*, 1027, doi:10.1029/2001GC000252, 2003.

Chan, T. F., G. H. Golub, and R. J. LeVeque, Algorithms for computing the sample variance: analysis and recommendations, *Amer. Statistician*, *37*, 242–247, 1983.

Clauset, A., C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, *SIAM Rev.*, *51*, 661–703, 2009.

Cleveland, W. S., *The Elements of Graphing Data*, Wadsworth, Monterey, 1985.

Cleveland, W. S., *Visualizing Data*, Hobart Press), Summit, N.J., 1993.

Constable, C., On rates of occurrence of geomagnetic reversals, *Phys. Earth Planet. Inter.*, *118*, 181–193, 2000.

Cox, D. R., and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, Methuen, London, 1966.

Cuite, C., N. Weinstein, K. Emmons, and G. Colditz, A test of numeric formats for communicating risk probabilities, *Med. Decis. Making*, *28*, 377–384, doi:10.1177/0272989X08315246, 2008.

Davison, A. C., and D. V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, New York, 1997.

Dekking, F. M., C. Kraaikamp, H. P. Lopuhaä, and L. E. Meester, *A Modern Introduction to Probability and Statistics: Understanding Why and How*, Springer, London, 2005.

Dym, H., and H. P. McKean, *Fourier Series and Integrals*, Academic Press, New York, 1972.

Efron, B., and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.

Ellsworth, W. L., The San Andreas Fault: earthquake history, 1769-1989, *U. S. Geol. Surv. Prof. Pap.*, *1515*, 1990.

Evans, M., N. Hastings, and J. B. Peacock, *Statistical Distributions*, Wiley, New York, 2000.

Fisher, N. I., *Statistical Analysis of Circular Data*, Cambridge University Press, New York, 1993.

Fisher, N. I., T. Lewis, and B. J. J. Embleton, *Statistical Analysis of Spherical Data*, Cambridge University Press, New York, 1987.

Gershenfeld, N. A., *The Nature of Mathematical Modeling*, Cambridge University Press, New York, 1999.

Gigerenzer, G., Z. Swijtink, T. Porter, L. Daston, J. Beatty, and L. Krüger, *The Empire of Chance: How Probability Changed Science and Everyday Life*, Cambridge University Press, New York, 1989.

Heaton, T. H., Tidal triggering of earthquakes, *Geophys. J. Roy. Astron. Soc.*, *43*, 307–326, 1975.

Heaton, T. H., The 1971 San Fernando earthquake; a double event?, *Bull. Seismol. Soc. Am.*, *72*, 2037–2062, 1982.

Hoffrage, U., S. Lindsey, R. Hertwig, and G. Gigerenzer, Communicating statistical information, *Science*, *290*, 2261–2262, 2000.

Hogg, R. V., and E. A. Tanis, *Probability and Statistical Inference*, Prentice-Hall, Upper Saddle River N.J., 2000.

James, F., RANLUX: A Fortran implementation of the high-quality pseudorandom number generator of Luscher, *Comput. Phys. Commun.*, *79*, 111–114, doi:10.1016/0010-4655(94)90233-X, 1994.

Kagan, Y. Y., Statistical aspects of Parkfield earthquake sequence and Parkfield prediction experiment, *Tectonophys.*, *270*, 207–219, 1997.

Lancaster, P., and K. Salkauskas, *Curve and Surface Fitting: an Introduction*, Academic Press, Orlando, 1986.

L'Ecuyer, P., and R. Simard, TestU01: A C library for empirical testing of random number generators, *ACM Trans. Math. Softw.*, *33*, 22, doi: 10.1145/1268776.1268777, 2007.

Lüscher, M., A portable high-quality random number generator for lattice field theory simulations, *Comput. Phys. Commun.*, *79*, 100–110, doi:10. 1016/0010-4655(94)90232-1, 1994.

Marsaglia, G., Choosing a point from the surface of a sphere, *Ann. Math. Statist.*, *43*, 645–646, 1972.

Matsumoto, M., and T. Nishimura, Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Trans. Modeling Comput. Simul.*, *8*, 3–30, 1998.

McFadden, P. L., Statistical tools for the analysis of geomagnetic reversal sequences, *J. Geophys. Res.*, *89*, 3363–3372, 1984.

McFadden, P. L., and R. Merrill, Lower mantle convection and geomagnetism, *J. Geophys. Res.*, *89*, 3354–3362, 1984.

Press, W. H., S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing, 2nd Ed.*, Cambridge University Press, Cambridge, 1992.

Rice, J. A., *Mathematical Statistics and Data Analysis*, Duxbury MA, Belmont, 1995.

Rice, J. A., *Mathematical Statistics and Data Analysis*, third ed., Duxbury MA, Duxbury Press, 2006.

Rutherford, E., and H. Geiger, The probability variations in the distribution of alpha-particles, *Phil. Mag. (Ser. 6)*, *20*, 698–707, 1910.

Salsburg, D., *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, W. H. Freeman, New York, 2001.

Seber, G. A. F., and A. J. Lee, *Linear Regression Analysis*, Wiley, New York, 2003.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1986.

Sivia, D. S., *Data Analysis: a Bayesian Tutorial*, Clarendon Press, Oxford, 1996.

Stigler, S. M., Stigler's Law of Eponymy, *Trans. New York Acad. Sci. Ser. 2*, *39*, 147–157, 1980.

Stigler, S. M., Terrestrial mass extinctions and galactic plane crossings, *Nature*, *313*, 159, 1985.

Stuart, A., and J. K. Ord, *Kendall's Advanced Theory of Statistics: Distribution Theory*, Edward Arnold, London, 1991.

Stuart, A., and J. K. Ord, *Kendall's Advanced Theory of Statistics: Classical Inference and Relationship*, Edward Arnold, London, 1994.

Tauxe, L., *Paleomagnetic Principles and Practice*, Kluwer, Dordrecht, 1998.

Thomas, D. B., W. Luk, P. H. W. Leong, and J. D. Villaseñor, Gaussian random number generators, *ACM Computing Surv.*, *39*, 11, 2007.

Tufte, E. R., *Envisaging Information*, Graphics Press, Cheshire, Conn., 1991.

Tufte, E. R., *Visual Explanations: Images And Quantities, Evidence And Narrative*, Graphics Press, Cheshire, Conn., bF241.T83, P93.5.T846, 1997.

Tufte, E. R., *The Visual Display of Quantitative Information (2nd ed.)*, Graphics Press, Cheshire, Conn., 2001.

Wall, J. V., and C. R. Jenkins, *Practical Statistics for Astronomers (2nd ed.)*, Cambridge University Press, New York, 2012.

Wells, D. L., and K. J. Coppersmith, New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seism. Soc. Am.*, *84*, 974–1002, 1994.

Wyatt, F. K., and D. C. Agnew, The PIN1 and PIN2 GPS Sites at Piñon Flat Observatory, *Sio technical report*, Scripps Institution of Oceanography, available at `http://escholarship.org/uc/sio_techreport`, 2004.