

List of references for data sets

(used by Dr. Laura J. Simon's Stat 501 courses at Penn State University)

NOTE: A special thanks to all of the following whose data sets have been used (at least once) in my graduate-level regression course (Stat 501) at Penn State University. I (and my students!) especially appreciate how these data sets have enriched the student's experience in learning how to apply the methods of regression analysis to "real-world" data.

A SECOND NOTE: Some of the data sets have multiple references – one from the original source and two (or more) from the secondary source (the place where I most likely found the data set.)

A THIRD NOTE: Every attempt has been made to report the references to the data sets completely and accurately. If you notice a problem with any reference – in error or in omission – please don't hesitate to notify me by e-mail at lsimon@stat.psu.edu.

American Automobile Association. (1991). *Defensive Driving: Managing Time and Space*, Pamphlet #3389.

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27, pp. 17-21. (**anscombe.txt**)

Atkinson, A.C. (1985). *Plots, Transformations, and Regression*. Oxford: Clarendon Press, p.4. (**alpswater.txt**)

Atkinson, A.C. (1994). Transforming both sides of a tree. *The American Statistician* 48: 307-312. (**shortleaf.txt**)

Bruce, R.A., Kusumi, F. and D. Hosmer. (1973). Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *American Heart Journal* 65: 546-562. (**treadmill.txt**)

Bruce, C. and F. X. Schumacher. (1935). *Forest Mensuration*. McGraw Hill: New York. (**shortleaf.txt**)

British official statistics, *Family Expenditure Survey*, Department of Employment, 1981. (**alcoholtobacco.txt**)

Chambers, John M., Cleveland, William S., Kleiner, Beat, and Tukey, Paul A. (1983). *Graphical Methods for Data Analysis*. Wadsworth: Belmont, CA. (**automobile.txt**, **hamster.txt**)

Colby, C., Kilgore, D. L., Jr. and S. Howe. (1987). Effects of hypoxia and hypercapnia on V_T , f and V_I of nestling and adult bank swallows. *American Journal of Physiology* 253: R854-R860. (**babybirds.txt**)

Cook, R. Dennis and Sanford Weisberg. (1999). *Applied Regression Including Computing and Graphics*. Wiley: New York. (**bluegills.txt**)

Criqui, M. H. University of California, San Diego, School of Medicine. *New York Times*, 28 Dec 1994. (**wineheart.txt**)

David, Sandra K. and William T. Riley. (1990). The relationship of the Allen Cognitive Level Test to cognitive abilities and psychopathology. *American Journal of Occupational Therapy* 44: 493-497. (**allentest.txt**)

Daniel, Wayne W. (1999). *Biostatistics: A Foundation for Analysis in the Health Sciences*. Wiley: New York. (**allergen.txt**, **allentest.txt**, **bloodpress.txt**, **stress.txt**, **birthsmokers.txt**, **depression.txt**, **dexterity.txt**, **surgerytemp.txt**)

Denby, L. and D. Pregibon. (1987). An example of the use of graphics in regression. *The American Statistician* 41: 33-38. (**oldfaithful.txt**)

De Veaux, Richard D. and Velleman, Paul F. (2004). *Intro Stats*. Pearson-Addison Wesley: Boston, MA. (**boyleslaw.txt**)

Domeck, Doug, Smithers Scientific Services, Inc., Akron, OH. (**treadwear.txt**)

Draper, Norman R. and Harry Smith. (1998). *Applied Regression Analysis*, 3rd edition. Wiley: New York. (**cement.txt**, **adaptive.txt**, **corrosion.txt**, **lifeline.txt**, **treesize.txt**)

Duncan, D. F. (1994). Drug law enforcement expenditures and drug-induced deaths. *Psychological Reports*, 75, 57-58. (**drugdea.txt**)

Erickson, Roberta S. and Sue T. Yount. (1991). Effect of aluminized covers on body temperature in patients having abdominal surgery. *Heart and Lung* 20: 255-264. (**surgerytemp.txt**)

Exploring Data (website). <http://www.exploringdata.cqu.edu.au/>. (**challenger.txt**, **alligator.txt**)

Ezekiel, M. and K. A. Fox. (1959). *Methods of Correlation and Regression Analysis*. Wiley: New York. (**cornyield.txt**)

Fisher, Lloyd D. and Gerald van Belle. (1993). *Biostatistics: A Methodology for the Health Sciences*. Wiley-Interscience: New York. (**skincancer.txt**, **treadmill.txt**)

Fordedal, H. et al. (1995). A multivariate analysis of W/O emulsions in high external electric fields as studied by means of dielectric time domain spectroscopy. *Journal of Colloid and Interface Science* 173(2): 398, Table 2. (**wateroil.txt**)

Glantz, Stanton A. and Bryan K. Slinker. (2001). *Applied Regression and Analysis of Variance*, 2nd edition. McGraw-Hill: New York. (**babybirds.txt**, **coolhearts.txt**)

Grayson, D. K. (1990). Donner party deaths: a demographic assessment. *Journal of Anthropological Research* 46: 223-42. (**donner.txt**)

Hald, A. (1952). *Statistical Theory with Engineering Applications*. Wiley: New York. (**cement.txt**)

Hale, S. L., Dave, R. H. and R. A. Kloner. (1997). Regional hypothermia reduces myocardial necrosis even when instituted after the onset of ischemia. *Basic Research in Cardiology* 92: 351-357. (**coolhearts.txt**)

Hamilton, Lawrence C. (1992). *Regression with Graphics: A Second Course in Applied Statistics*. Duxbury Press: Belmont, CA. (**leadcord.txt**)

Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J., and Ostrowski, E, Eds. (1994). *A Handbook of Small Data Sets*. Chapman & Hall: London. (**alpswater.txt**, **usair.txt**, **wordrecall.txt**)

Hughes, Susan. School of Public Health, University of Illinois, Chicago. Courtesy use of data set. (**hospital.txt**)

- Iman, Ronald L. (1995). *A Data-Based Approach to Statistics, Concise Version*. Duxbury: Belmont, CA. (**alcoholarm.txt**)
- Karelitz, S. Fisichelli, V. R., Costa, J., Kavelitz, R. and L. Rosenfeld. (1964). Relation of crying in early infancy to speech and intellectual development at age three years. *Child Development* 35: 769-777. (**cryingiq.txt**)
- Keienburg, W. Heinemann, D, and S. Schmitz, eds. (1990). *Grizmek's Encyclopedia of Mammals*. McGraw-Hill: New York. (**mammgest.txt**)
- Kleinbaum, David G., Kupper, Lawrence, L., Muller, Keith E., and Azhar Nizam. (1998). *Applied Regression Analysis and Other Multivariate Methods*. Duxbury Press: Pacific Grove: CA. (**delinquency.txt**)
- Last Resource, Inc. Bellefonte, PA. (**seeingdist.txt**)
- Lyman, C. P., O'Brien, R. C., Greene, G. C., and Papafrangos, E. D. (1981). Hibernation and longevity in the Turkish hamster *Mesocricetus brandti*. *Science* 212, 668-670. (**hamster.txt**)
- Margolin, Barry H. (1988). Statistical aspects of using biological markers. *Statistical Science* 3(3): 351-357. (**urine.txt**)
- Marsh, C. (1988). *Exploring data*. Polity Press: Cambridge, UK. (**husbandwife.txt**)
- McClave, J. T. and F. H. Dietrich, II. (1994). *Statistics*, 6th edition. Dellen-MacMillan: New York. (**entrance.txt**)
- McEvoy ,P. and C. Cox. (1991). Successful biological control of ragwort, *Senecio jacobaea*, by introduced insects in Oregon. *Ecological Applications* 1(4): 430-42. (**pestcontrol.txt**)
- Mendenhall, William and Terry Sincich. (2003). *Regression Analysis: A Second Course in Statistics*. Pearson Prentice Hall: Upper Saddle River, NJ. (**ojsweet.txt, tirepressure.txt, whitespruce.txt, wateroil.txt**)
- Mendenhall, W.M, Parsons, J. T., Stringer, S. P., Cassissi, N. J. and R. R. Milion. (1989). T2 oral tongue carcinoma treated with radiotherapy: Analysis of local control and complications. *Radiotherapy and Oncology* 16: 275-282. (**tongue.txt**)
- Mickey, M. R., Dunn, O. J. and V. Clark. (1967). Note on the use of stepwise regression in detecting outliers. *Computers and Biomedical Research* 1: 105-111. (**adaptive.txt**)
- Misner, E. G. Studies of the relationship of weather to production and price of farm products, I. Corn, Cornell University, March 1928. (**cornyield.txt**)
- Moore, David S. and George P. McCabe. (1993). *Introduction to the Practice of Statistics*, 2nd edition. Freeman: New York. (**adaptive.txt, alcoholtobacco.txt**)
- Moore, David S. (1997). *Statistics: Concepts and Controversies*, 4th edition. Freeman: New York. (**wineheart.txt**)
- Mosteller, F., Rourke, R. E. K., and G. B. Thomas. (1970). *Probability with Statistical Applications*, 2nd edition, 383, Table 11-1. Addison Wesley: Reading, Massachusetts. (**wordrecall.txt**)
- Myers, Raymond H. (1986). *Classical and Modern Regression with Applications*. Duxbury Press: Boston, MA. (**oxygen.txt, oxygentrain.txt, oxygentest.txt**)

Neter, J., Kutner, M., Nachtsheim, C. and W. Wasserman. (1996). *Applied Linear Regression Models*, 3rd edition. Irwin: Chicago. (**alphapluto.txt**, **newaccounts.txt**)

Rabinowitz, Michael, Needleman, Herbert, Burley, Michael, Finch, Hollister, and John Rees. (1984). Lead in umbilical blood, indoor air, tap water, and gasoline in Boston. *Archives of Environmental Health* 39(4): 299-301. (**leadcord.txt**)

Ramsey, Fred L. and Daniel W. Schafer. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis*. Duxbury Thomson Learning: Pacific Grove, CA. (**pestcontrol.txt**, **sexdiscrim.txt**, **donner.txt**, **cornyield.txt**, **bldgstories.txt**)

Roberts, H. V. "Harris Trust and Savings Bank: An Analysis of Employee Compensation." (1979). Report 7946, Center for Mathematical Studies in Business and Economics, U. of Chicago Graduate School of Business. (**sexdiscrim.txt**)

SAS Institute, Inc., *SAS User's Guide: Statistics*, 1982 edition. SAS Institute: Cary, NC. (**oxygen.txt**, **oxygentrain.txt**, **oxygenest.txt**)

Scholz, H. Northern Lights College, British Columbia. (**whitespruce.txt**)

Sokal, R. R. and F. J. Rohlf. (1981). *Biometry*, 2nd edition. W. H. Freeman: San Francisco. (**usair.txt**)

Smith, Gary. (1998). *Introduction to Statistical Reasoning*. McGraw-Hill: Boston, MA. (**adaptive.txt**, **cryingiq.txt**)

Tamhane, Ajit C. and Dorothy D. Dunlop. (2000). *Statistics and Data Analysis from Elementary to Intermediate*. Prentice Hall: Upper Saddle River, NJ. (**alpswater.txt**, **anscombe.txt**, **cement.txt**, **iqsize.txt**, **oldfaithful.txt**, **treadwear.txt**, **hospital.txt**, **entrance.txt**, **wordrecall.txt**, **mammgest.txt**, **tongue.txt**)

Tanner, M. (1996). *Tools for Statistical Inference*, 3rd edition. Springer: New York, p. 28. (**tongue.txt**)

The 1994 World Almanac. 1993. Funk and Wagnalls: Mahwah, NJ. (**bldgstories.txt**)

Urbano-Marquez, A., et al. (1989). The effects of alcoholism on skeletal and cardiac muscle. *The New England Journal of Medicine*, Vol. 320, No. 7, 409-415. (**alcoholarm.txt**)

U. S. Cancer Mortality by County: 1950-1959 [1974]. DHEW Publication No. (NIH) 74-615, Bethesda, MD. (**skincancer.txt**)

U. S. Census Bureau. <http://www.census.gov>. (**uspopn.txt**)

Utts, Jessica M. (1999) *Seeing through Statistics*, second edition. Duxbury: Pacific Grove, CA. (**bookprice.txt**)

Utts, Jessica M and Robert F. Heckard. (2004). *Mind on Statistics*, second edition. Duxbury Brooks/Cole: Belmont, CA. (**heightspeed.txt**, **seeingdist.txt**, **earthquake.txt**, **carstopping.txt**)

Weiss, Neil A. (2002). *Introductory Statistics*, 6th edition. Addison Wesley: Boston. (**shortleaf.txt**)

Wild, Christopher J. and George A. F. Seber. (2000). *Chance Encounters: A First Course in Data Analysis and Inference*. Wiley: New York. (**infant.txt**, **drugdea.txt**)

Willerman, L., Schultz, R., Rutledge, J. N., and E. Bigler. (1991). In vivo brain size and intelligence. *Intelligence* 15: 223-228. (**iqsize.txt**)

Wilson, M. E. and L. E. Mather. (1974). Letter to the editor. *Journal of the American Medical Association* 229(11): 1421-1422. (**lifeline.txt**)

Witteman, Agnes, M., Stapel, Steven O., Perdok, Gerrard J., Sjamsoedin, Deman H.S., Jansen, Henk M., Aalberse, Rob C., and Jaring S. van der See. (1996). The relationship between RAST and skin test results in patients with asthma or rhinitis: A quantitative study with purified major allergens. *Journal of Allergy and Clinical Immunology* 97: 16-25. (**allergen.txt**)

Woods, H., Steinour, H. H. and H. R. Starke. (1932). Effect of composition of Portland cement heat evolved during hardening. *Industrial and Engineering Chemistry* 24: 1207-1214, Table I. (**cement.txt**)

Zar, Jerrold H. (1999). *Biostatistical Analysis*, fourth edition. Prentice Hall: Upper Saddle River, NJ. (**electricfish.txt**)