



## StatNews #21

### Data Sets on the Web April 1997

Suppose you are teaching a class and need a small data set to illustrate a concept or to analyze for a class project. Maybe you already know of useful data from a textbook but do not want to type it in by hand, or you would like to re-analyze data you read about in a journal article. Thanks to the availability of web-based data archives, the solution to your data needs may be close at hand. While data sets are available in many places on the web, two sites, DASL and StatLib, are especially comprehensive. The Data and Story Library (DASL, pronounced "dazzle", at <http://lib.stat.cmu.edu/DASL/>) contains many small data sets that are useful for illustrating concepts in a class or for conducting short data analysis projects. Data sets are indexed by topic, and are currently available for almost 30 topics from agriculture to zoology. For example, four data sets are listed under the topic "Nutrition", including a data set containing the sodium and calorie content of different brands of hot dogs. DASL's search engine allows you to find data sets by topic (health, social science, politics, etc.), by statistical method (regression, t-test, etc.), or both. For example, a search for "regression" and "biology" yielded data on mortality of fruit flies and on mercury contamination in bass. DASL includes a "story" for each data set that provides background on the data and suggests ways to analyze the data. The story associated with mercury contamination in bass described the way the data were collected and suggested transformations of the variables that might clarify some of the relationships. StatLib (<http://lib.stat.cmu.edu/>) is an archive for statistical computer routines, data sets, and other information. Look here for data from more than 15 textbooks including Case Studies in Biometry (Lange et al, 1994), and Categorical Data Analysis (Agresti, 1990), and from several academic journals. While the list of data sets is extensive, the search function is limited. The best way to find data in this archive is to read through the list of data sets names and descriptions. StatLib and DASL provide data in text format that can be downloaded and read into any statistical software package. If you need help with this process, contact the Office of Statistical Consulting. If you are aware of other web sites that contain data archives, please let us know.

Author: Cara Olsen

(This newsletter was distributed by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at [cscu@cornell.edu](mailto:cscu@cornell.edu). Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at <http://www.cscu.cornell.edu>).