APPENDIX C

Descriptions of Data Sets

CONTENTS

C.1 FLORIDA LAKE DATA

To obtain this data in EXCEL or SAS data format, go to the text Web site at http://www.elsevierdirect.com and download DATATAB_LAKES.xls or DATATAB_LAKES.sas7bdat.

This data was obtained from the Web site of Florida Lakewatch (http://lakewatch.ifas.ufl.edu/), a volunteer organization coordinated through the University of Florida's Institute of Food and Agricultural Sciences, Fisheries and Aquatic Sciences. The organization aims to provide scientifically collected data that can be used to understand and manage the state's lakes. We thank the organization for their permission to cite this data.

Total chlorophyll is a measure of free-floating algae in water. Typically, the amount of algae is limited by the amount of nutrients in the water and by seasonal influences such as temperature. The primary limiting nutrients in Florida are usually either phosphorous, nitrogen, or some combination. For more information on the meaning of the variables and the role of limiting nutrients, see Florida Lakewatch's Circular #102, *A Beginner's Guide to Water Management: Nutrients*, available on the Web site listed earlier.

Statistical Methods, DOI: 10.1016/B978-0-12-374970-3.00023-8

This data is from the 2005 data report, and covers lakes in the north central counties of Alachua, Bradford, Clay, and Putnam. Each lake had measurements taken on a monthly basis. The following measurements are recorded for a cold-weather month (preferably January, but if that was missing then February or December) and a hot-weather month (preferably August, but if that was missing then July or September).

Variables	
Missing values for quantitative variables are coded as blanks in the Excel file and as missing values (.) in the SAS data file.	
Lake	Name of lake (character)
County	Name of county (character)
Geol_form	Name of the dominant geologic formation (see notes) HAW Hawthorne formation BON Bone Valley formation HOL Holocene JAX Jacksonville Limestone PLE Pleistocene PLI Pliocene
Soil	Name of the dominant soil type CS = clayey sand PS = phosphatic sand QS = quartzite sand OTHER = many other types where there were only one or two examples of each in the data set
Wtrmonth	Name of month for which the winter data is reported
WtrTP	Winter value for total phosphorous, in μg/L
WtrTN	Winter value for total nitrogen, in μg/L
WtrChlo	Winter value for total chlorophyll, in μg/L
WtrSecc	Winter value for Secchi depth, in feet
Smrmonth	Name of month for which the summer data is reported
SmrTP	Summer value for total phosphorous, in μg/L
SmrTN	Summer value for total nitrogen, in μg/L
SmrChlo	Summer value for chlorophyll, in μg/L
SmrSecc	Summer value for Secchi depth, in feet

Notes:

There are a total of 64 lakes in the data set, but data is missing for some variables.

Secchi depth is a measure of water clarity, with higher values denoting greater clarity. A Secchi disk is a disk with black and white wedges of color. It is lowered into the water, and the depth at which the wedges disappear is the Secchi depth.

Geologic formation. Two lakes actually overlapped formations. Newnan is both Bone Valley and Hawthorne. Chipco is both Hawthorne and Pleistocene.

C.2 STATE EDUCATION DATA SET

To obtain this data in EXCEL or SAS data format, go to the text Web site at http://www.elsevierdirect.com, and download DATATAB_EDUC.xls or DATATAB_EDUC.sas7bdat.

A version of this data was much discussed by conservative columnists in the mid 1990s, who saw that a regression of total SAT scores on per capita expenditures had a negative slope. They used this to argue that government expenditures were actually counterproductive. This argument was quickly refuted when others pointed out that it is necessary to control for the percentage of a state's students who are taking the SAT.

This is a newer version of the same data, taken mostly from the National Center for Education Statistics (U.S. Department of Education) Web site. Along with percentage taking the exam and per capita expenditures, we added information on state median income, state poverty rate, and state mean score on one component of the National Assessment of Educational Progress (NAEP) test.

This data is both observational and aggregated. That is, we do not have values of SAT scores for individual students, but summaries (aggregations) of information across many students.

For each state and the District of Columbia, the following values are recorded.

State	Character Value
SATcrit	Average score on the SAT critical reading test for all those taking the exam during the 2005/2006 year. National Center for Education Statistics,[1] Table 137.
SATmath	Average score on the SAT math test for all those taking the exam during the 2005/2006 year. National Center for Education Statistics, Table 137.
SATTotal	Average total score, sum of critical reading and math averages. National Center for Education Statistics, Table 137.
TakePCT	Percent of high school seniors taking the SAT. National Center for Education Statistics, Table 137.
Expend_pc	State per capita expenditures on instruction in elementary/secondary schools, FY2005. National Center for Education Statistics, Table 6.
NAEP_math8	Average scale score for mathematics among eighth graders in 2005 on the NAEP. National Center for Education Statistics, Table 129.

(*Continued*)