Stat 215b (Spring 2004): Lab 3

B. M. Bolstad
bolstad@stat.berkeley.edu

Due Mar 11, 2004 by 2:30pm

Introduction

Millions of Americans are routinely exposed to ozone in the ambient air at levels that are known to cause adverse health effects. While there is a great deal of knowledge about outdoor ozone concentrations, little is known about indoor concentrations and even less about personal exposures. Measurements of personal exposures are important in that estimates based on ambient concentrations alone may result in substantial misclassification of the exposure status of study subjects. A small passive ozone sampler has made possible assessment of personal ozone exposures in large field studies.

In this lab, we will consider data from a pilot study conducted during the summer of 1991 in State College, Pennsylvania. Extensive indoor, outdoor, and personal ozone measurements were collected for 19 children using passive ozone samplers. Detailed time-activity information was also collected for these children. These data were used to validate the new passive ozone sampler and to identify factors that affect personal ozone exposures. The primary goal was to develop a multiple regression exposure model to predict personal exposure in the children.

The passive ozone sampler consists of a badge clip supporting a barrel-shaped device. The limit of detection for the passive ozone sampler was 17.5 parts per billion (ppb) for 12-hour measurements. The passive ozone sampler was used at indoor and outdoor home sites, at a stationary site, and to measure exposures of individual children.

For validation purposes, outdoor ozone concentrations at a stationary site were measured continuously by the U.S. Environmental Protection Agency designated ozone analyzer. The limit of detection of this method is 2 ppb with a precision of 2 ppb.

Data

Ozone concentrations were measured from July 8 through August 27, 1991. Ozone samples were collected on days exhibiting a wide range of ozone concentrations. Indoor, outdoor, and personal samples were collected for 19 children (ages 10 and 11), all living in non-smoking households in one of six residential regions. Monitoring was conducted at each child's home for up to a six-day period. Up to three children were monitored each period. Regions 1 (downtown), 2, 4, and 5 are densely populated, while regions 3 and 6 are less populated communities, having been only recently developed.

Outdoor ozone concentrations were measured at the State College National Dry Depository Network site approximately 6 km west of downtown State College. At this stationary site, 12-hour average samples were collected twice daily (8 am to 8 pm and 8 pm to 8 am) using both passive ozone samplers and continuously using a photometric ambient ozone analyzer. Continuous monitoring also allows collection of 1-hour average measurements. At each home, indoor samples were collected over 12 hours for both daytime (8 am to 8 pm) and nighttime (8 pm to 8 am) periods using passive samplers. These samplers were placed in the main activity room of the child's home, at least 1 meter away from walls, windows, and air conditioners or other ventilation devices, and 1.2 meters above the floor. Outdoor ozone concentrations were measured using passive samplers placed outside homes, at least 1 meter from walls, trees, or other large objects. Outdoor samples were collected for 24-hour periods beginning at 8 am.

Personal exposures were measured during the day (8 am to 8 pm) using passive samplers. Samplers were pinned to the strap of a backpack worn by each participant throughout the monitoring period. Each participant also recorded his/her activities in a notebook during daytime monitoring periods. These entries were later aggregated into half-hour periods and were transferred onto formatted time-activity sheets by field technicians.

Before proceeding further, we first establish notation. The response variable is the 12-hour average daytime personal ozone concentration (Y) for participants on different days. The covariate values (X) have the subscript "1" for measured ozone concentrations, or "2" for data extracted from the time-activity diaries. The superscripts are mnemonic, with "D" for daytime, "N" for nighttime, "C" for continuous sampling, "P" for passive sampling, and "S" for staying near the home.

The data is located on the webpage. There are three files. The file `ozoneA.dat` contains the validation data collected at the stationary ambient monitoring site. The first column gives the date; the next four columns give X_1^{DC}, X_1^{DP}, X_1^{NC}, and X_1^{NP}. These are the 12-hour average daytime continuous, daytime passive, nighttime continuous, and nighttime passive samples, respectively. The file `ozoneB.dat` contains the personal ozone exposure data. The columns give the subject number, the date, the home region (regions 1 to 6), the response variable (Y), the 12-hour average daytime personal ozone concentration, the daytime and nighttime continuous ozone concentrations at the stationary site (X_1^{DC} and X_1^{NC}), the 24-hour average outdoor ozone concentration (X_1^O), the 12-hour average daytime and nighttime indoor ozone concentrations (X_1^{DI} and X_1^{NI}), the fraction of time spent anywhere outdoors (X_2^O), the fraction of time spent at home indoors (X_2^I), and finally, whether or not the child stayed near the home for the entire day ($X_2^S = 1$ when yes and 0 when no). The last three variables were extracted from the time-activity diaries. Note that it is possible to have $X_2^O + X_2^I < 1$ when a child spends a portion of a day in an outdoor environment other than their home (e.g. shopping mall, friend's home). Note that the ozone measurements are in ppb, and that `NA` denotes missing data.

Your Job

There are two major issues that you want to address in your report:

1. Is the passive sampler reliable? Test this by comparing the measurements of the passive sampler and the EPA-designated ozone analyzer at the stationary site.

2. Develop a linear model for predicting personal ozone exposure. You want the model to be

justifiable, i.e. you want to be able to give a reasonable explanation of why this model is appropriate. You may use forward selection in combination with backwards deletion (this is known as the stepwise procedure). Also, consider using AIC or BIC. As always, evaluate your fitted model by looking at residuals etc.

In addition, in the file `ozoneC.dat` there are a set of 19 additional observations made during a follow-up study. Use your model to predict personal ozone exposure for these 19 cases. How much error would you expect there to be in your predictions?