Homework 3

You may work in groups.
This homework is due at the beginning of class Thursday, February 5, 2009.
If you submit this homework by e-mail please send it to both me
(Judith.K.Dunn@uth.tmc.edu) and Yi-Ju (Yi-Ju.Chiang@uth.tmc.edu)

The dataset is HW3old.dta

Y’all have gotten a little lax about turning your homework in on time. I have heard every excuse in the books. I understand there are special circumstances but you should notify me or one of the TA’s prior to such problems. Otherwise we expect the homework be to turned in on time. Please!!!

The data for this homework is taken from The Statistical Sleuth by Fred Ramsey and Daniel Schafer (Duxbury,2002). Ramsey and Schafer based it on data from “For Debate: Pet birds as an Independent Risk Factor for Lung Cancer” by P.A. Holst, D. Kromhout and R. Brand (British Medical Journal, 1988).

A 1972-1981 health survey in The Hague, Netherlands discovered an association between keeping pet birds and increased risk of lung cancer. To investigate bird-keeping as a risk factor, researchers conducted a case control study of patients in 1985 at 4 hospitals in The Hague (population 450,000). They identified 49 cases of lung cancer among patients who were registered with a general practice, who were age 65 or younger, and who had resided in the city since 1965. They then selected 98 controls from the population of residents having the same general age structure.

What Ramsey and Schafer wanted you to do was to use logistic regression to consider bird-keeping as a risk factor for lung cancer while controlling for known risk factors such as years of smoking and socioeconomic status.

We, of course, aren’t quite ready for that yet. However, we can consider some of the preliminary steps for such analyses.

Prior to launching into logistic regression we need to know what variables are related to the outcome variable (here lung cancer). So I want to you find out if there is an association between lung cancer and each of bird-keeping, socioeconomic status, and years of smoking.

This is a dataset we have not used before so you first need to check for missing data and check the distributions of the variables for the four variables of interest.

Then do the appropriate analyses and describe your answers in a way that would be suitable for inclusion in a grant proposal.