Statistics 850 **yandell** **Spring 1995**

Midterm—Due Wed 22 Mar

The midterm has two interrelated problems. Prepare a written report geared toward a statistician with the background of Stat 850. Use a combination of plots (diagnostic, interaction, other?) and small summary tables to supplement each report. Label plots and tables clearly. Your report should not exceed 10 pages (no smaller than 10 point size characters!), and can be shorter. Typing is not required, but neatness is!

A nutrition student conducted a survey of nutrition and health among 5th and 8th grade students (roughly 10- and 13-year-old boys and girls). They were asked to complete an 8-page survey. Only a portion of that material is considered here. One of the main questions for the researcher concerned the relationship between knowledge of nutrition (`knowscor`) and health beliefs about the effects of fat. The first examination of the data considers only whites as they comprised about 90% of the students.

An overall measure of nutrition knowledge, `knowscor`, was developed as the number of correct answers on 13 questions. Most of these questions concerned the amount of fat or calories in food items. Health beliefs about fat (`fatbel`) was the sum of three questions about the eventual effect of having a lot of fat in one's food.

1. The researcher would like to be able to explain fat beliefs (`fatbel`) in terms of `gender` and `grade`, as modified by knowledge (`knowscor`). In other words, is gender (or grade) important? Does knowledge score appear to alter fat beliefs, and does this differ by gender and/or grade? Be sure to justify your steps.

2. The nutritional scientist is a bit uneasy about aggregating three fat belief questions into one. Now that you have worked out some methods for analysis, examine the separate responses concerning belief of the effect of eating fat on heart disease (`hdfat`), getting fat (`fatfat`) and cancer (`canfat`). Your final models may be different for these three.

3. Did you get the same factors and interactions in the two problems? Explain briefly why the separate models in problem 2 might be preferred to the model for the aggregated response in problem 1. Alternatively, argue for the aggregated response model. Make your points concise. You may include statistical as well as nutritional points.

The full data set, containing about 1200 students, can be found in `mida.dat`. A 20% subset, with about 250 students, is in `midb.dat`. You may use either one.