Biostatistics 208 **Winter 2010**

HOMEWORK 2:
Due Tuesday, February 9, 2010

Please hand in hardcopy in lecture

For this homework use the dataset `hwk2.dta`, available on the course web site. The data are a subset of the data from HERS, the trial of hormone therapy for secondary prevention of coronary heart disease. The dataset includes the following variables for the 2,032 diabetes-free women with glucose $\leq$125 mg/dl:

- ht (treatment assignment: 1 = hormone therapy, 0 = placebo)
- glucose0 (fasting glucose at baseline, in mg/dl)
- glucchange (change in glucose from baseline to yr 1 in mg/dl)
- exercise (at least 3 times per week: 1 = yes, 0 = no)
- age (in years at baseline)
- raceth (race/ethnicity: 1 = white, 2 = African American, 3 = other)
- smoker (current smoker at baseline: 1 = yes, 0 = no)
- bmi (body mass index at baseline, in kg/m^2)

Please prepare a Word file with answers to each of the following questions. Paste relevant STATA output into the file to illustrate your points. In particular, the regression output on which you base your conclusions should be included. Using the `Courier` font for the STATA output keeps it properly aligned, and setting the font size to 8 saves paper.

This part of the exercise is motivated by the Diabetes Prevention Project, which showed a protective effect of exercise and weight loss in preventing progression to diabetes among women with fasting glucose levels between 100 and 125 mg/dl.

1. (1 pt.) What is the crude or unadjusted association of regular exercise with fasting glucose at baseline?
2. (2 pt.) Use a multi-predictor linear model to estimate the adjusted association of exercise with `glucose0`, adjusting for age, race/ethnicity, and smoking. Is there substantial confounding of the association between exercise and baseline glucose by the other factors? Is exercise a significant predictor after adjustment?

3. (1 pt.) Suppose we were considering including BMI in the adjusted model for Question 2, but were unsure whether it should be seen as a confounder or mediator of the association between exercise and glucose levels. In practice, it can be difficult to distinguish confounding from mediation, and is not possible on statistical grounds. Instead, it is judgment call based on plausible models of causation. Do you think BMI is more plausibly a confounder or a mediator of the association between exercise and glucose levels? Describe a plausible mechanism for the effect of exercise on glucose under which BMI is a mediator. Come up with one under which BMI is a confounder. Which do you think is more credible?

4. (2 pt.) A previous HERS paper has shown that hormone therapy prevented progression to diabetes. Does change in glucose from baseline to the first annual visit (`glucchange`) differ in the hormone therapy and placebo groups? What is the average change in each group?

5. (4 pt.) It is not uncommon for the effects of a treatment on a surrogate outcome like glucose levels to depend on the baseline level of the surrogate. Check for interaction between hormone therapy and baseline glucose level in a model for change in glucose between baseline and year 1. Give and interpret the statistical test of interaction. Regardless of the result of the interaction test result, use the interaction model to estimate the effect of hormone therapy on glucose change (`glucchange`), with a 95% CI, among patients with

   - baseline glucose of 80 mg/dL
   - baseline glucose of 90 mg/dL
   - baseline glucose of 100 mg/dL
   - baseline glucose of 110 mg/dL

   Comment on the pattern in stratum-specific estimates, in light of the result of the test for interaction.

6. (1 pt.) Extra Credit: Repeat the previous problem with centered baseline glucose. Interpret the coefficient estimate for `ht`.

7. (1 pt.) Extra Credit: Produce a single graph for the fitted regression model showing the association of baseline glucose with change in glucose for each of the treatment groups.