Exercise 4LHT: Using PC SAS to identify the factors influencing height (HT) of women in Hawaii.

Objectives:

1. To apply the procedures learned in class and Exercise 3 to the analysis of experimental data.
2. To determine the "best" model to predict height (HT) of women in Hawaii.

Background:

The purpose of this study conducted by Dr. R. Novotny (Dept. of Human Nutrition, Food and Animal Science, CTAHR) was to identify factors related to height and leg length of women in Hawaii. The data were drawn from a project investigating life style factors related to bone mass (Dr. J. Davis, Hawaii Osteoporosis Center, PI). Data were obtained from a retrospective questionnaire concerning diet, menstruation, physical activity, socio-demographic factors and current height and sitting height measurements of 180 women ages 25-35. Height is a key measurement used to assess past nutritional status. Height is also commonly used in conjunction with other measures, such as weight, to assess current nutritional state. Also, variations in body proportions (here, the difference in leg length among women) offer further information as to the timing of growth spurts. Longer leg length, for example, suggests a later adolescent growth spurt.

Women in Hawaii are of varying ethnic backgrounds and, therefore, vary in food consumption patterns, physical activity and other lifestyle characteristics while living in the same climate and with similar infectious disease exposure.

Subjects and methods:

Volunteer women of each of four ethnic groups (Caucasian, Japanese, Filipino, Hawaiian/part-Hawaiian) were solicited via posters and pamphlets, TV and radio announcements. Subjects were required to have both parents of the same ethnic group, except Filipinos and Hawaiians who had to be at least one-eighth or more Filipino and Hawaiian, respectively. Subjects were also required to be between 25 and 35 years of age.

Milk consumption was recorded using a scale of one to four, where one was three or more glasses of milk per day, two was one to two glasses per day, three was every week but not daily, and four was rarely or never. Three age periods were examined for milk consumption (10-14 years, 15-19 years, and 20-24 years).

Procedure:

1. The data set is in an Excel file on the class lab web site.

2. The variables and their order in the data set are indicated on the attached sheet which is a partial printout of the entire data set. You will notice that in the partial data
3. The dependent variables, HT and LEGLENGTH are listed first and are given in cm. You will use HT as the dependent variable in your models. BMI stands for Body Mass Index and is calculated as Wt/Ht^2. The ethnic groups Japanese, Caucasian, Filipino and Hawaiian are listed next. The AGE of the woman is given next in years and the woman's age when she had her first menstrual period (MENARCHE) follows. Milk consumption at the three ages mentioned above are presented next and the last variable is the woman's age at birth in pounds.

4. You are to use HT as the dependent variable in your models. You may want to create some square, square root, or log terms of some of the variables to see if this improves the prediction. Some examples are Age^2, Square Root(Age), and Log(Age). These terms can be created using Data Transform in SAS Analyst.

5. You can add other variables that you believe are needed to give you the analysis that you want.

6. You can use Regression and Stepwise or other options to develop your models. Since HT and LEGLENGTH are both dependent variables, they should not be used as independent variables.

7. Next Wednesday please turn in printouts of the following items for the models that you consider as potential models for this relationship.

   1. Analysis of variance and parameter estimates
   2. Plot of the residuals for the final selected model
   3. A summary of the regression statistics (SSE, MSE, Rsq, Adj-Rsq, PRESS)
   4. The prediction equation for the model you consider to be the "best" model
   5. An explanation why you selected that particular model over the other models.

References:
4. Draper and Smith 1981. Regression Analysis

These data were donated by Dr. J. Davis to provide a learning experience for this class. They are not to be used for other purposes or published, without permission from Dr. Davis.

H.M. Zaleski/J.A. Silva/J.L. Brewbaker