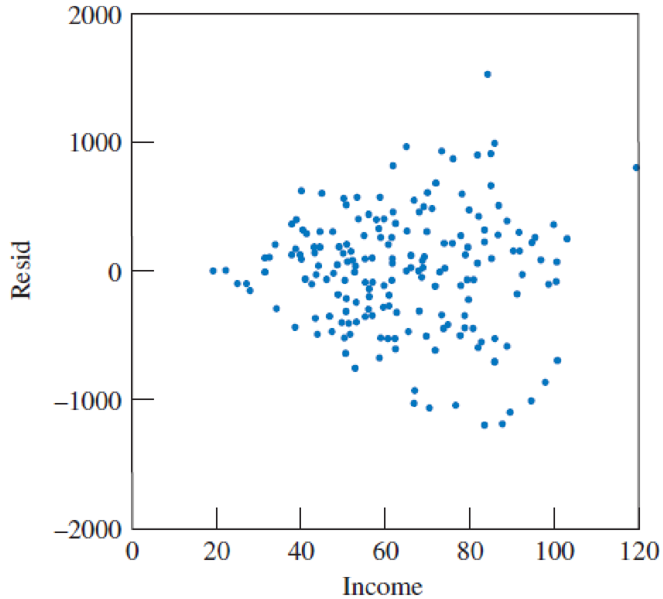
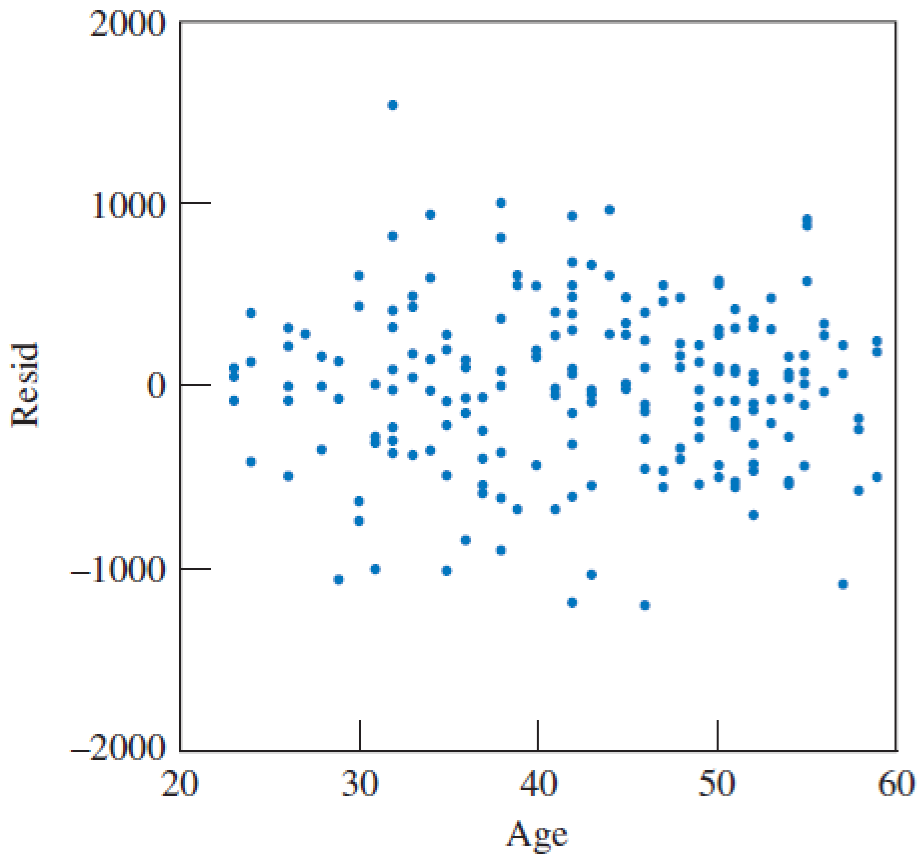
8.4 A sample of 200 Chicago households was taken to investigate how far American households tend to travel when they take vacation. Measuring distance in miles per year, the following model was estimated

$$MILES = \beta_1 + \beta_2 INCOME + \beta_3 AGE + \beta_4 KIDS + e$$





4 Residual plots for Exercise 8.4: vacation data.

The variables are self-explanatory except perhaps for *AGE*, the average age of the adult members of the household. The data are in the file *vacation.dat*.

(a) The equation was estimated by least squares and the residuals are plotted against age and income in Figure 8.4. What do these graphs suggest to you?

(b) Ordering the observations according to descending values of *INCOME*, and applying least squares to the first 100 observations, and again to the second 100 observations, yields the sums of squared errors

$$SSE_1 = 2.9471 \times 10^7 \quad SSE_2 = 1.0479 \times 10^7$$

Use the Goldfeld–Quandt test to test for heteroskedastic errors. Include specification of the null and alternative hypotheses.

(c) Table 8.2 contains three sets of estimates: those from least squares, those from least squares with White's standard errors, and those from generalized least squares under the assumption $\sigma_i^2 = \sigma^2 \times INCOME^2$.

   (i) How do vacation miles traveled depend on income, age, and the number of kids in the household?

   (ii) How do White's standard errors compare with the least squares standard errors? Do they change your assessment of the precision of estimation?

   (iii) Is there evidence to suggest the generalized least squares estimates are better estimates?