(iii) What is the predicted value of *GPA* when $ACT = 20$?

(iv) How much of the variation in *GPA* for these eight students is explained by *ACT*? Explain.

2.3 Let *kids* denote the number of children ever born to a woman, and let *educ* denote years of education for the woman. A simple model relating fertility to years of education is

$$kids = \beta_0 + \beta_1 educ + u,$$

where u is the unobserved error.

(i) What kinds of factors are contained in u? Are these likely to be correlated with level of education?

(ii) Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

2.4 Suppose you are interested in estimating the effect of hours spent in an SAT preparation course (*hours*) on total SAT score (*sat*). The population is all college-bound high school seniors for a particular year.

(i) Suppose you are given a grant to run a controlled experiment. Explain how you would structure the experiment in order to estimate the causal effect of *hours* on *sat*.

(ii) Consider the more realistic case where students choose how much time to spend in a preparation course, and you can only randomly sample *sat* and *hours* from the population. Write the population model as

$$sat = \beta_0 + \beta_1 hours = u$$

where, as usual in a model with an intercept, we can assume $E(u) = 0$. List at least two factors contained in u. Are these likely to have positive or negative correlation with *hours*?

(iii) In the equation from part (ii), what should be the sign of β_1 if the preparation course is effective?

(iv) In the equation from part (ii), what is the interpretation of β_0?

2.5 Consider the savings function

$$sav = \beta_0 + \beta_1 inc + u,\ u = \sqrt{inc}\cdot e,$$

where e is a random variable with $E(e) = 0$ and $Var(e) = \sigma_e^2$. Assume that e is independent of *inc*.

(i) Show that $E(u|inc) = 0$, so that the key zero conditional mean assumption (Assumption SLR.4) is satisfied. [*Hint:* If e is independent of *inc*, then $E(e|inc) = E(e)$.]

(ii) Show that $Var(u|inc) = \sigma_e^2 inc$, so that the homoskedasticity Assumption SLR.5 is violated. In particular, the variance of *sav* increases with *inc*. [*Hint:* $Var(e|inc) = Var(e)$, if e and *inc* are independent.]

(iii) Provide a discussion that supports the assumption that the variance of savings increases with family income.

2.6 Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the OLS intercept and slope estimators, respectively, and let $\bar{u}$ be the sample average of the errors (not the residuals!).

(i) Show that $\hat{\beta}_1$ can be written as $\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i$ where $w_i = d_i/SST_x$ and $d_i = x_i - \bar{x}$.

(ii) Use part (i), along with $\sum_{i=1}^{n} w_i = 0$, to show that $\hat{\beta}_1$ and $\bar{u}$ are uncorrelated. [*Hint:* You are being asked to show that $E[(\hat{\beta}_1 - \beta_1)\cdot \bar{u}] = 0$.]