A multiple regression fit results in

$$y = 19.9 - 1.79x_1 + .0432x_2 + 0.556x_3 + 1.11x_4 - 1.79x_5$$

Is the policy implication that it is best to pay teachers small salaries and not educate mothers? Clearly, many of the predictors are highly correlated with each other and are also correlated with variables that are not in the model, and literal interpretation of a coefficient as being the effect if that variable is increased by one unit and the others are held fixed is fallacious. Also note that this is an observational study, not a controlled experiment.

14.6 Conditional Inference, Unconditional Inference, and the Bootstrap

The results in this chapter on the statistical properties of least squares estimates have been derived under the assumptions of a linear model relating independent variables $\mathbf{X}$ to dependent variables $\mathbf{Y}$ of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

In this formulation, the independent variables have been assumed to be *fixed* with randomness arising only through the errors **e.** This model seems appropriate for some experimental setups, such as that of Section 14.1, where fixed percentages of dies, **X,** were used and peak areas on a chromatograph, **Y,** were measured. However, consider Example B of Section 14.2.2, where the flow rate of a stream was related to its depth. The data consisted of measurements from 10 streams and it would seem to be rather forced to model the depths of those streams as being fixed and the flow rates as being random. In this section, we pursue the consequences of a model in which both $\mathbf{X}$ and $\mathbf{Y}$ are random, and we discuss the use of the bootstrap to quantify the uncertainty in parameter estimates under such a model.

First we need to develop some notation. The design matrix will be denoted as a random matrix $\boldsymbol{\Xi}$ and a particular realization of this random matrix will be denoted, as before, by $\mathbf{X}$**.** The rows of $\boldsymbol{\Xi}$ will be denoted by $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n$ and the rows of a realization $\mathbf{X}$ by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. In place of the model $Y_i = \mathbf{x}_i\boldsymbol{\beta} + e_i$, where $\mathbf{x}_i$ is fixed and e_i is random with mean 0 and variance σ^2, we will use the model $E(Y|\boldsymbol{\xi} = \mathbf{x}) = \mathbf{x}\beta$ and $\text{Var}(Y|\boldsymbol{\xi} = \mathbf{x}) = \sigma^2$. In the fixed $\mathbf{X}$ model, the e_i were independent of each other. In the random $\mathbf{X}$ model, Y and $\boldsymbol{\xi}$ have a joint distribution (for which the conditional distribution of Y given $\boldsymbol{\xi}$ has mean and variance as specified before) and the data are modeled as n independent random vectors, $(Y_1, \boldsymbol{\xi}_1), (Y_2, \boldsymbol{\xi}_2), \ldots, (Y_n, \boldsymbol{\xi}_n)$ drawn from that joint distribution. The previous model is seen to be a *conditional* version of the new model—the analysis is conditional on the observed values $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

We will now deduce some of the consequences for least squares parameter estimation under the new, unconditional, model. First, we have seen that in the old model, the least squares estimate of $\boldsymbol{\beta}$ is unbiased (Theorem A of Section 14.2.3); viewed within the context of the new model we would express this result as $E(\hat{\boldsymbol{\beta}}|\boldsymbol{\Xi} = \mathbf{X}) = \boldsymbol{\beta}$**.**