HLM

Hierarchical Linear Modeling

Lesson One

Lesson One Plan

I. Simple Linear Regression (one group)

   a) Model and assumptions
   b) Parameter estimates and standard errors
   c) Predicted values
   d) Centering

Lesson One Plan

II. Simple Linear Regression (two groups)

III. Simple Linear Regression (several groups)

   a) Motivation underlying HLM
   b) Example

I. Simple Linear Regression (one group)

We begin by specifying a theoretical model about how the world operates. This is our model which we base on some theory, prior research, and our beliefs and values. The simple linear regression model specifies a linear relationship between two constructs:

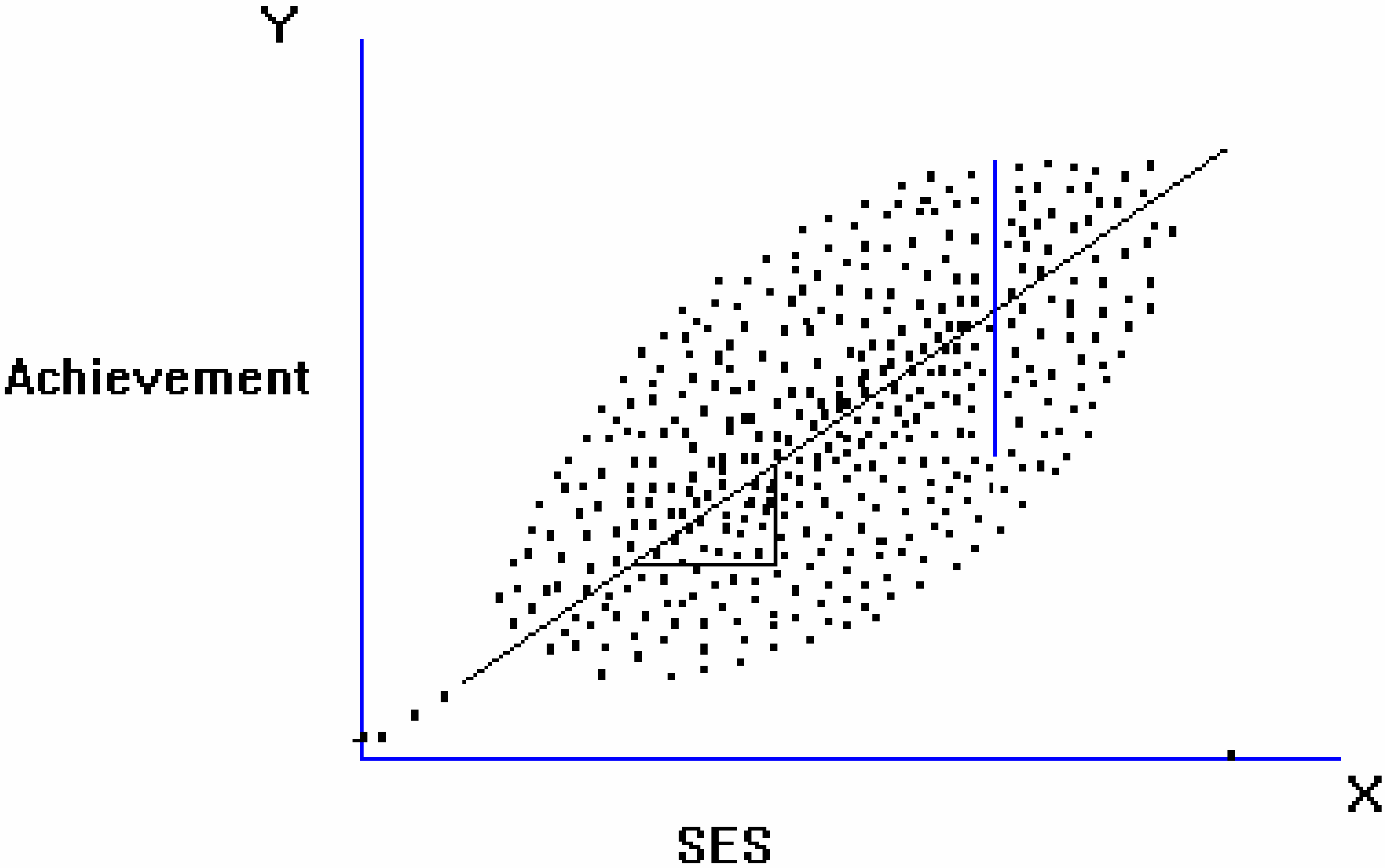
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_I$$

The model assumes that the residuals are normally distributed, independent, with a mean of zero and a constant variance, σ^2.

$$\varepsilon_i \sim NID(0, \sigma^2)$$

We collect data that are indicators of our constructs Y and X, and "fit" the data to the model. Using some estimation technique we can obtain estimates of our regression parameters:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$


Y
Achievement
X
SES

Presume that the data we collected are based on a sample from some larger population. Our exercise has only yielded *estimates* of our "true" population values. We want to know how precise our estimates are. Imagine repeating the exercise a number of times, each time recording the estimates of β_0 and β_1. The variances of our recorded β_0's and β_1's would indicate the precision of our estimates. The standard deviation of the recorded estimates are the standard errors.

In practice the standard errors are estimated empirically from the data. The size of the standard errors depends not only on how well the data fit the model, but also on the size of the sample and the heterogeneity of the X variable. Standard errors are smaller, indicating greater precision, when sample sizes are large and the variance of X is large.

Given the estimates of our parameters we can obtain a "predicted value" of Y for a particular value of X. The predicted value $\hat{Y}_0$ for the value X_0 is given by:

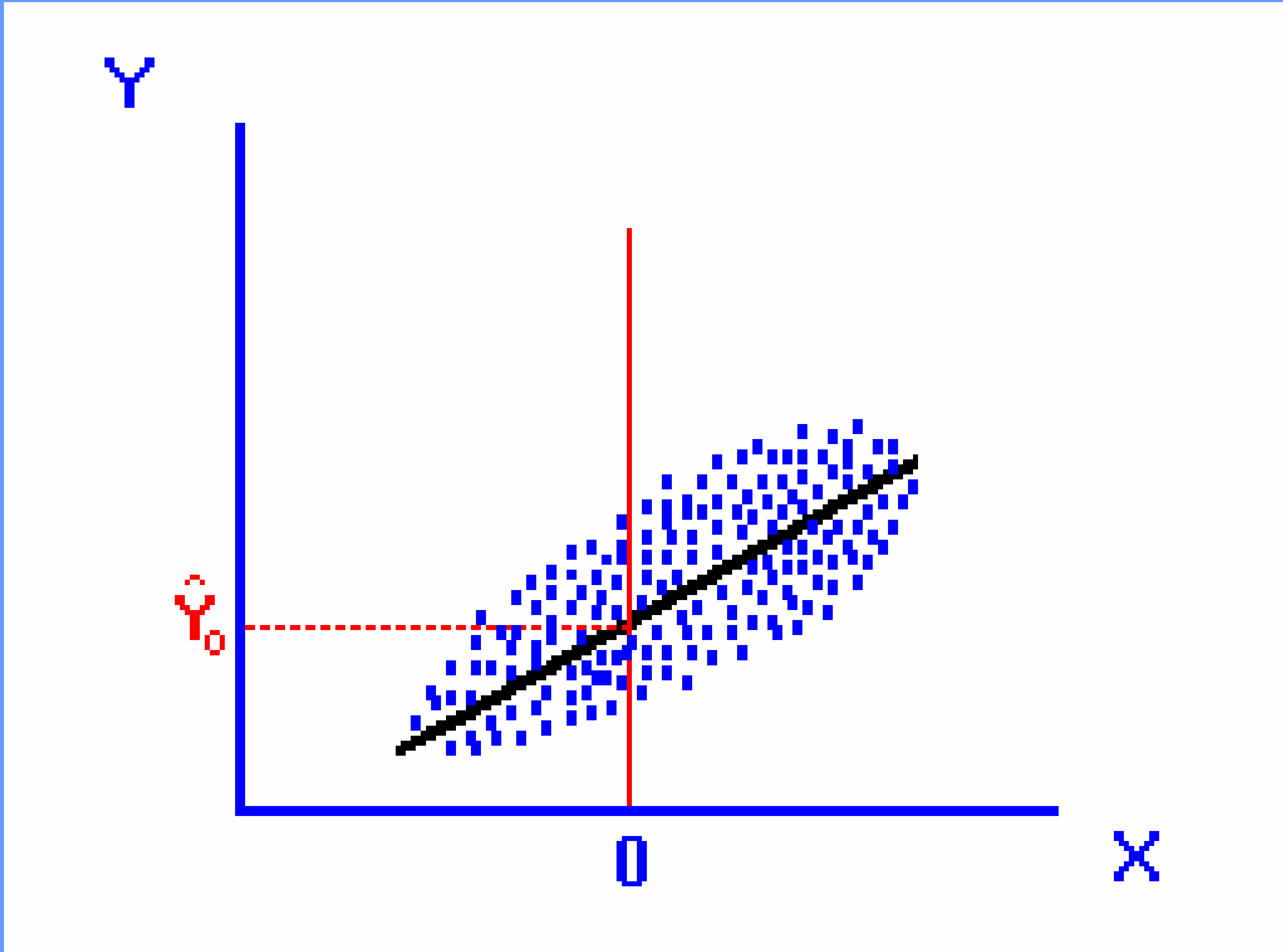
$$\hat{Y}_0 = \beta_0 + \beta_1 X_0$$

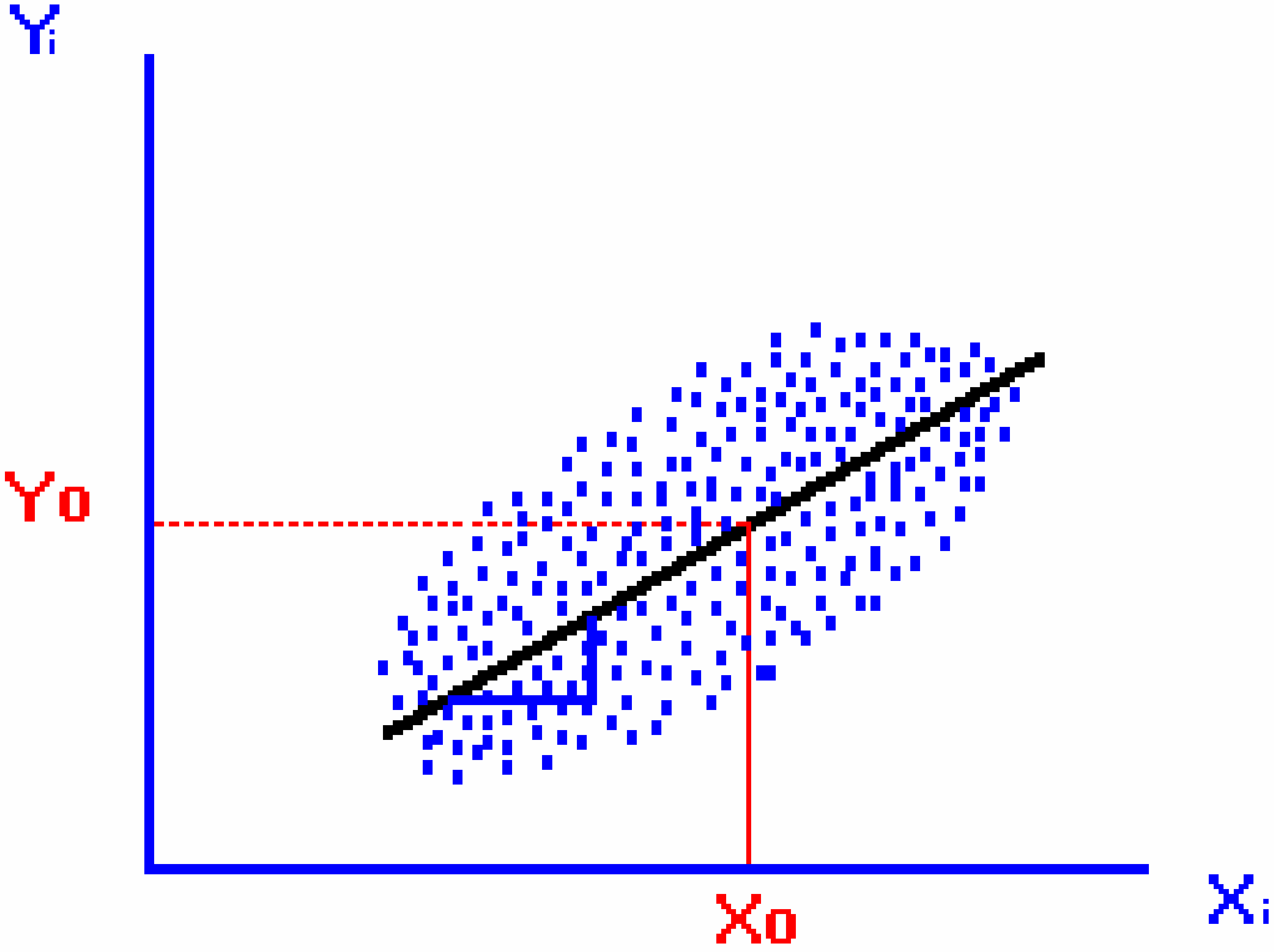
We can change the "location" of X values, but not their "scale" without changing the regression slope, β_1. In other words we can subtract a constant from every X value without affecting β_1.

$$Y_i = \beta_0' + \beta_1'(X_i - c) + \epsilon_i$$

A convenient transformation for multilevel analysis is to centre data on the grand mean for the sample, or if it is known, the mean for the population:

$$Y_i = \beta_0 + \beta_1'(X_i - \bar{X}) + \epsilon_i$$


Y
$\hat{Y}_0$
0
X


Yi
Yo
Xo
Xi

With the data "centred", the predicted value for a person who is average on X is simply β_0 :

$$\hat{Y}_0 = \beta_0 + \beta_1(X_0 - \bar{X}) = \beta_0 \quad \text{if } X_0 = \bar{X}$$

II. Simple Linear Regression (two groups)

Suppose that we have data describing two groups, such as females and males, a treatment and a control group, or simple data describing pupils in two separate schools. Usually we are interested in estimating the "sex difference", the "treatment effect", or the "school effect". These can be thought of as the difference in the predicted values for two groups for a particular value, X_0.

$$Y_{01} = \beta_{01} + \beta_{11}X_0$$

$$Y_{02} = \beta_{02} + \beta_{12}X_0$$

$$Y_{01} - Y_{02} = (\beta_{01} - \beta_{02}) + (\beta_{11} - \beta_{12})X_0$$

Notice that the estimate of the effect is $\beta_{01} - \beta_{02}$ at the value $X_0 = 0$, or if $\beta_{11} = \beta_{12}$. In general, however, the effect depends on what value of X we are considering.

In this case the treatment effect can be positive or negative (or zero) depending on the value of X one is considering.

In an analysis of covariance, (ANCOVA), we represent group membership with a dummy variable, which enables us to combine the two regression equations into a single analysis:

$Y_i = \beta_0 + \beta_1 X_i + \beta_2 G_i + \varepsilon_i$ where $G_i = 0$ for group 1

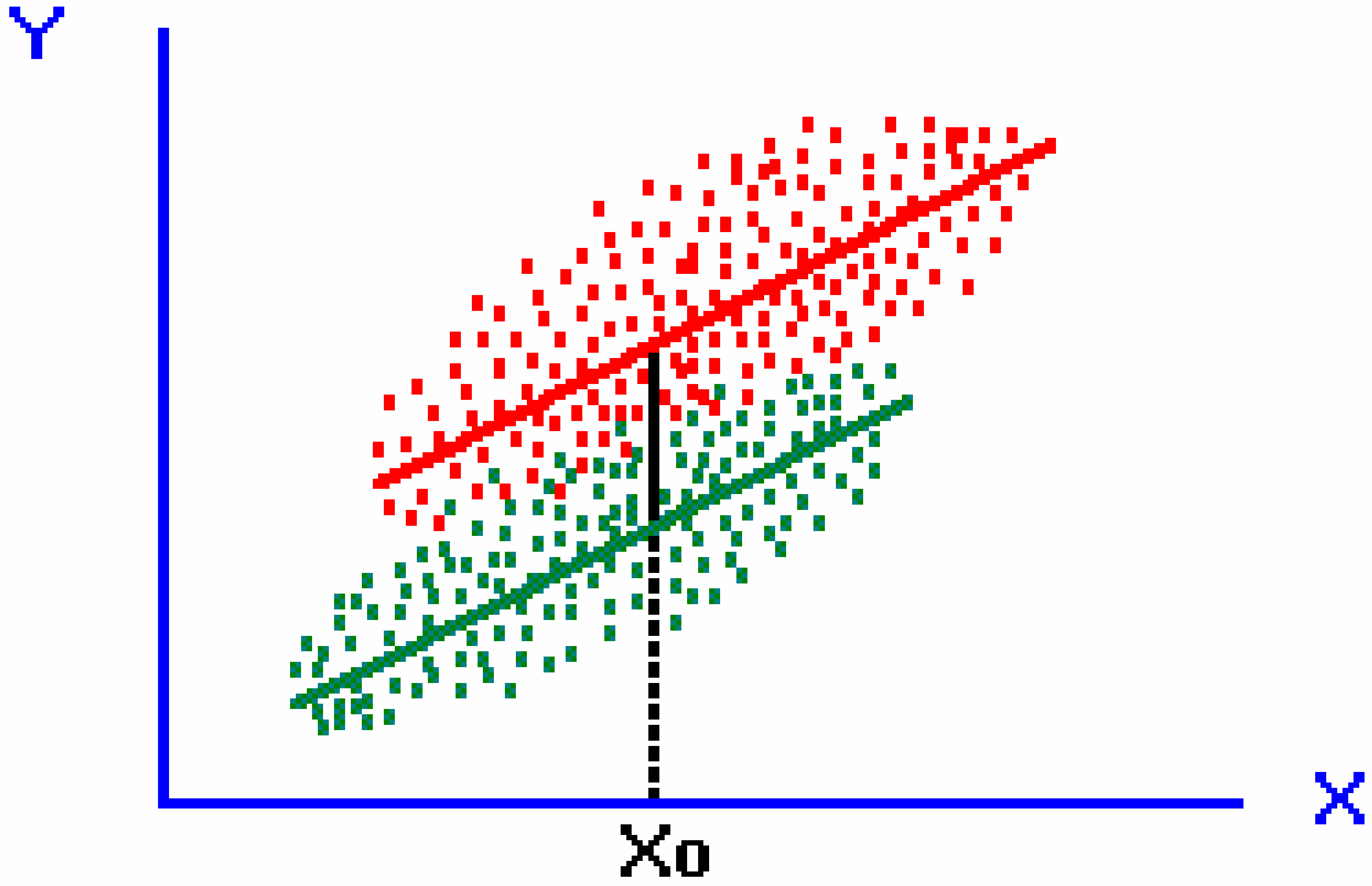
$G_i = 1$ for group 2

The predicted values for the two groups at X_0 are given by:

$Y_{01} = \beta_0 + \beta_1 X_0$ for $G_i = 0$

$Y_{02} = \beta_0 + \beta_1 X_0 + \beta_2$ for $G_i = 1$

The treatment effect, $Y_{01}-Y_{02}$, is given simply by $-\beta_2$. Thus the coefficient β_2 represents the effect on the outcome measure of membership in the group denoted 1.



Y
X
Xo

In the model above, there is only one slope, β_{1j}: that is, the slopes are constrained to be parallel for the two groups. We can model separate slopes for the two groups by including an interaction term:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 G_i + \beta_3 G_i X_i + \varepsilon_i$$

Now our predicted values are:

$$Y_{01} = \beta_0 + \beta_1 X_0$$

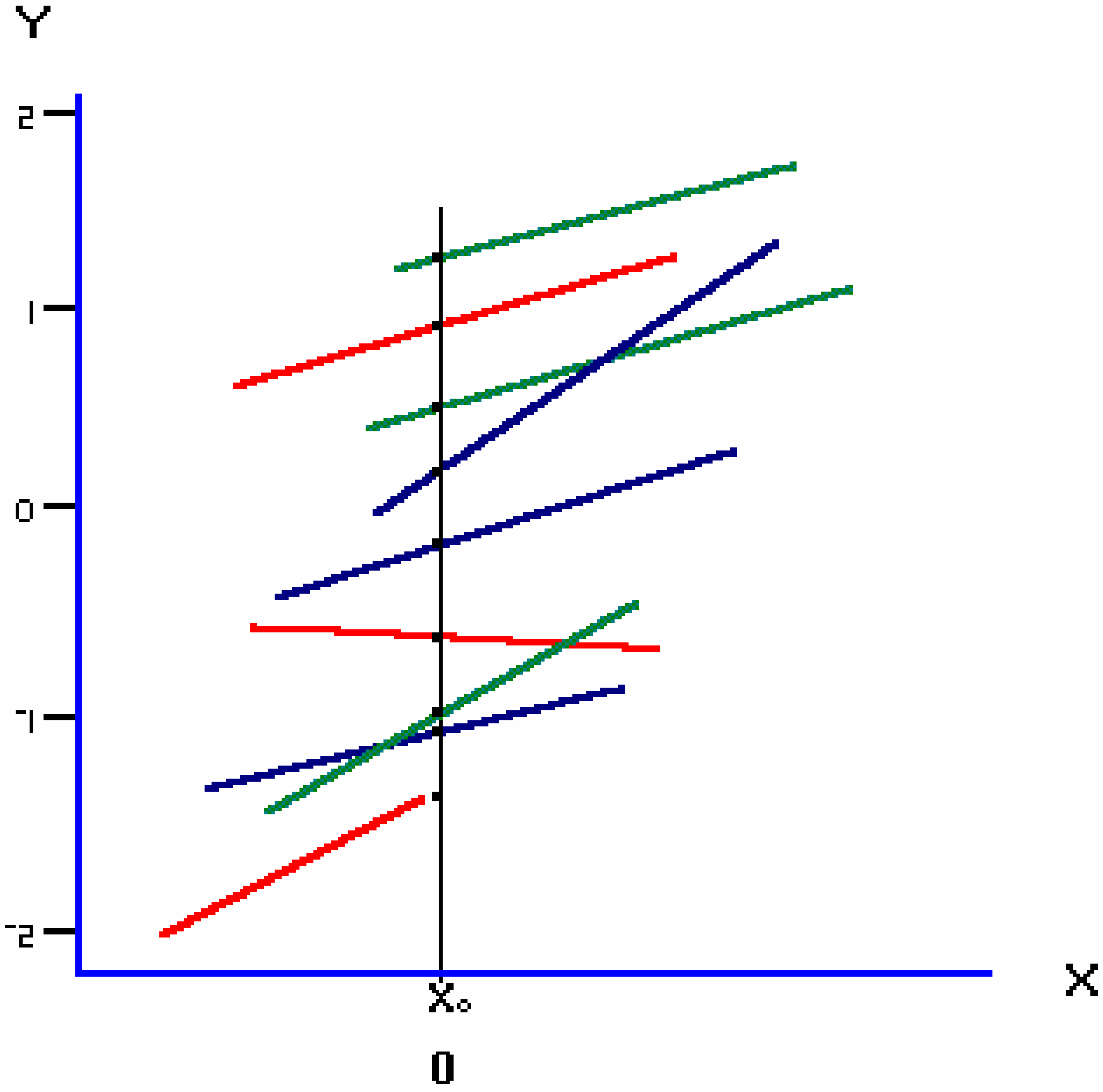
$$Y_{02} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_0$$

and the treatment effect is given by: $-\beta_2 - \beta_3 X_0$

In this case the regression slopes are non-parallel and the treatment effect depends on X_0.

Generally we would estimate the non-parallel slopes model first, and determine whether the interaction term was statistically significant. If it were not statistically significant, we would drop the interaction term and re-estimate the regression parameters. Notice, however, that in the centre of the data, the two models will yield similar results (identical results under some conditions).

We would prefer to use the parallel-slopes model because it is simpler, and if in reality the slopes are parallel, it yields more precise results. But if the slopes are truly non-parallel, and we constrain them to be parallel, we will obtain biased estimates of the treatment effect.


Y
2
1
0
¯1
¯2
X
X₀
0

III. Simple Linear Regression (Several Groups)

Now consider the case when we have data describing three or more groups. We can write separate equations for each group:

$$Y_{i1} = \beta_{01} + \beta_{11}X_{i1} + \varepsilon_{i1}$$

$$Y_{i2} = \beta_{02} + \beta_{12}X_{i2} + \varepsilon_{i2}$$

. . . .

. . . .

. . . .

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$$

Now the expected value of Y at $X_0 = 0$ is given by β_{0j}. Thus we can estimate a "treatment effect" for each of the J groups. If we centre Y on its grand mean, then the J estimates of treatment effects take on positive or negative values, which, as in analysis of variance, are estimates of the treatment effect relative to the grand mean.

Often we are interested in whether the treatment effects depend on some group-level factor, such as expenditures per pupil. We are also interested in how much the treatment effects vary.

Sometimes we are interested also in whether the slopes vary significantly, and if so, whether this variation is related to some group-level factor.

A multilevel data analysis enables us to:

a) estimate the effects of group-level factors on both the treatment effects and slopes.

b) estimate the "true" variance in the regression parameters across groups.

Later, when we consider the multivariate case (several X variables) we will be able to accomplish the two aims above, with control for confounding variables at both the individual and group levels.