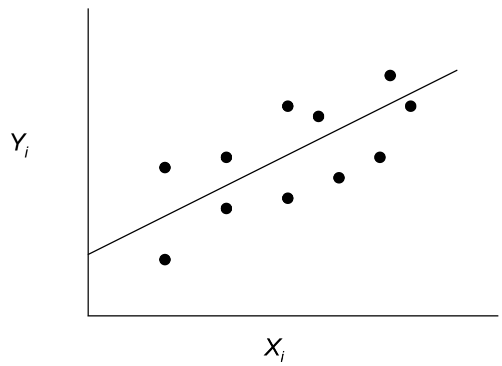
Regression Review

Regression analysis summarizes the relationship between two variables. The dependent variable, Y_i, is predicted by one or more independent variables, X_i. The subscript i, sometimes left off for convenience, represents an index for an individual score. For the first case, $i=1$, for the second case $i=2$ and so on. The relationship between X_i and Y_i is summarized by a line of best fit, nearest to the points on a scatterplot as possible. The equation for the regression line can be written as $\hat{Y}_i = \beta_0 + \beta_1 X_i$, where $\hat{Y}$ is the predicted value of Y, β_0 is the intercept, or value of Y when X = 0, and β_1 is the slope of the line. The slope represents the amount that Y changes for each increment of 1 in X. (Note that our text uses β for *unstandardized* regression coefficients)



The line does not fit perfectly, and there is some error in prediction of observed values of Y. Although many textbooks use e_i for errors or residuals, I will use notation R_i. Thus, the observed values of Y_i depend on the intercept, the value of X_i, the slope (the relation between X and Y), and error.

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

The intercept is not given much thought in most applications of regression, but it has a more important role in multilevel regression models. One can calculate the value of the intercept from the means of X and Y and the slope:

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

Considering this equation, the intercept can be thought of as highly related to the mean of Y. When there is no X for prediction, the mean of X is 0, or the slope is 0, the value of the intercept is actually equal to the mean of Y, $\beta_0 = \bar{Y}$. In all other cases, the intercept is the mean of Y "corrected" or "adjusted" by the amount $\beta_1 \bar{X}$ and, therefore, can be considered an adjusted mean.

Consider then a version of the regression equation with no X:

$$Y_i = \beta_0 + R_i$$

The value of an observed value of Y_i is equal to the intercept plus some error. Considering the intercept as the mean, this equation states that any observed value of Y is equal to the mean of Y plus some deviation from the mean of Y.