Definition

Let $(x_{1i}, x_{2i}, \cdots, x_{Ki}, y_i)$ for $i = 1, 2, \cdots, n$ be the observations of K independent (explanatory) variables x and one dependent variable y on a set s of n elements. Let also

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1K} \\ 1 & x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The hyperplane that minimizes the SSE is given by the least squares regression, $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K$ with $[b_0 \quad b_1 \quad \cdots \quad b_K]' = b$ and

$$b = (X'X)^{-1} X'Y. \qquad \square$$

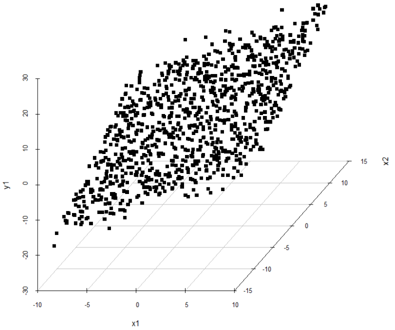

Figure: 3D Scatter plots of three variables x_1, x_2 and y.

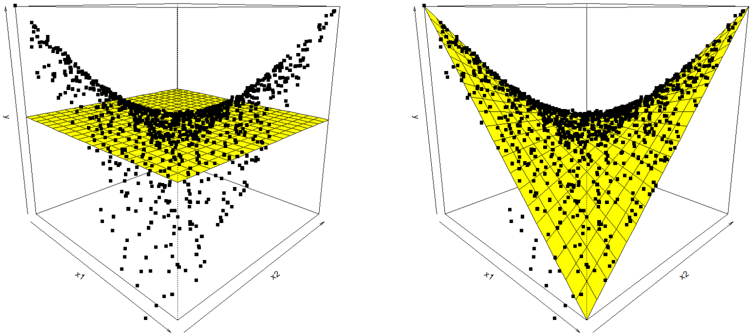

Figure: 3D Scatter plots and fitted regressions of three variables x_1, x_2 and y. Without interaction (left panel) and with interaction (right panel).

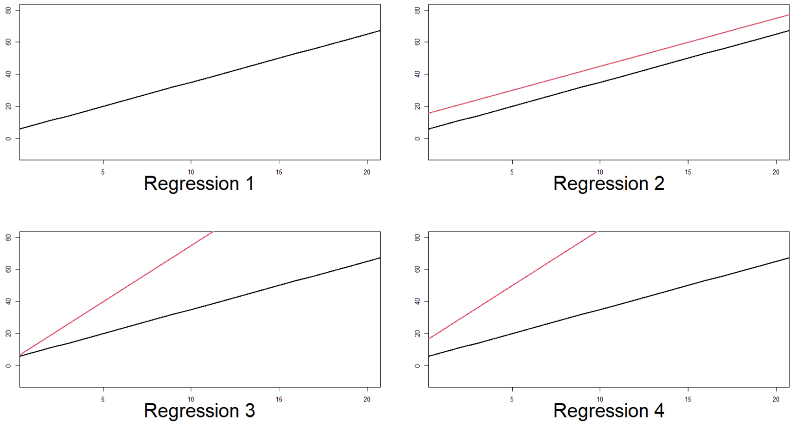

Figure: Plots of four regressions

Figure: Scatter plot of SAT scores vs. GPA of 4137 students. Red: athletes; Black: non-athletes

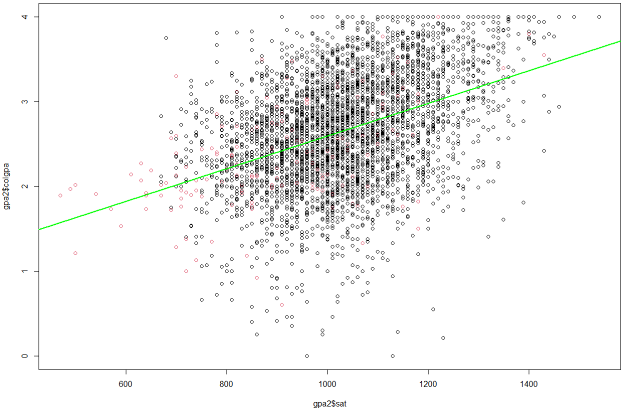

Figure: Scatter plot of SAT scores vs. GPA of 4137 students. Red: athletes; Black: non-athletes

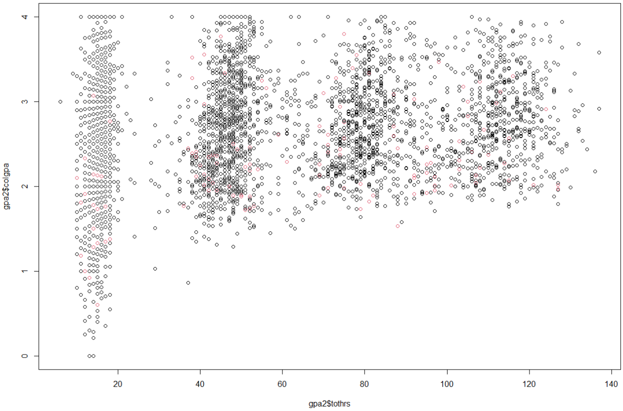

Figure: Scatter plot of *tothrs* vs. GPA of 4137 students. Red dots: athletes; Black dots: non-athletes

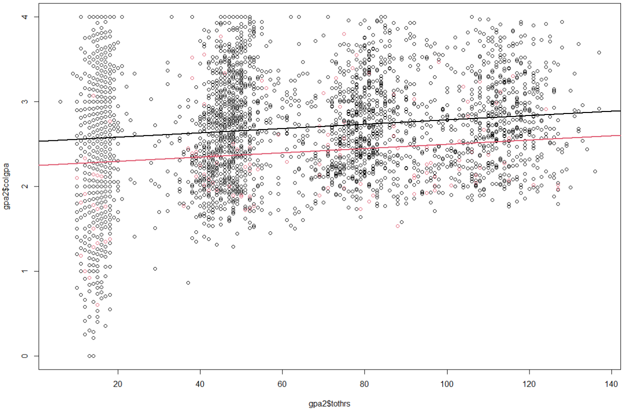

Figure: Scatter plot of *tothrs* vs. GPA of 4137 students. Red dots: athletes; Black dots: non-athletes

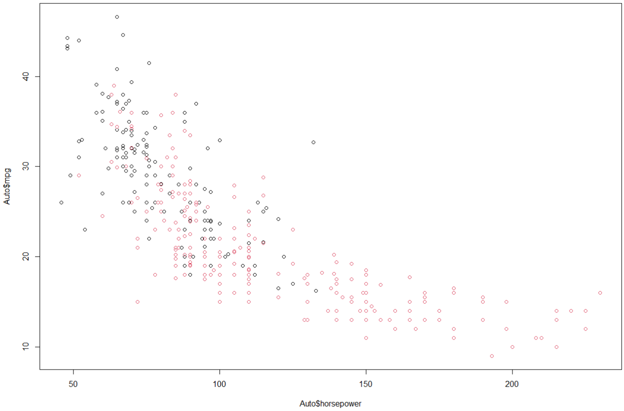

Figure: Scatter plot of horsepower vs. mpg of 392 automobiles. Red dots: American; Black dots: non-American

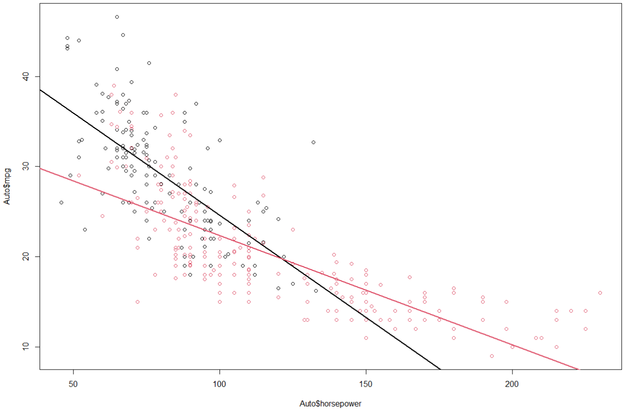

Figure: Scatter plot of horsepower vs. mpg of 392 automobiles. Red dots: American; Black dots: non-American

Definition (Coefficient of determination)

The *coefficient of determination* of a regression, denoted by R^2, is

$$R^2 = \frac{SSR}{SST} \qquad \text{or equivalently} \qquad R^2 = 1 - \frac{SSE}{SST}. \qquad \square$$

Definition (Adjusted Coefficient of determination)

The *adjusted coefficient of determination* of a multiple linear regression, denoted by $\bar{R}^2$, is

$$\bar{R}^2 = 1 - \frac{SSE/(n-K-1)}{SST/(n-1)}. \qquad \square$$