How Well Can We Predict?

To measure how well we can predict a response variable, y, from an explanatory variable, x, we can look at the average squared residual, which the least squares line tries to minimize.

Since the mean of the residuals is zero, this is the same as the variance of the residuals. The *residual standard deviation*, often called s, is the square root of this.

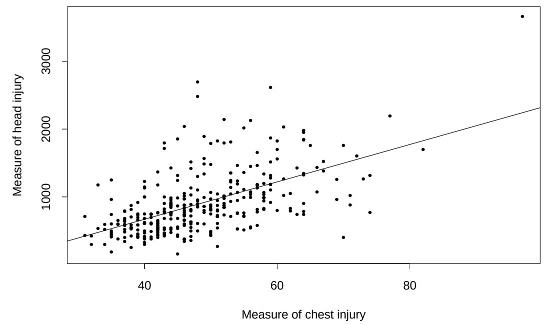
Another measure of predictability is r^2 (the *coefficient of determination*), which is the fraction of the variance of y that is "explained" by the regression model:

$$r^2 \;=\; \frac{\text{Var}(\hat{y})}{\text{Var}(y)} \;=\; 1 - \frac{s^2}{\text{Var}(y)}$$

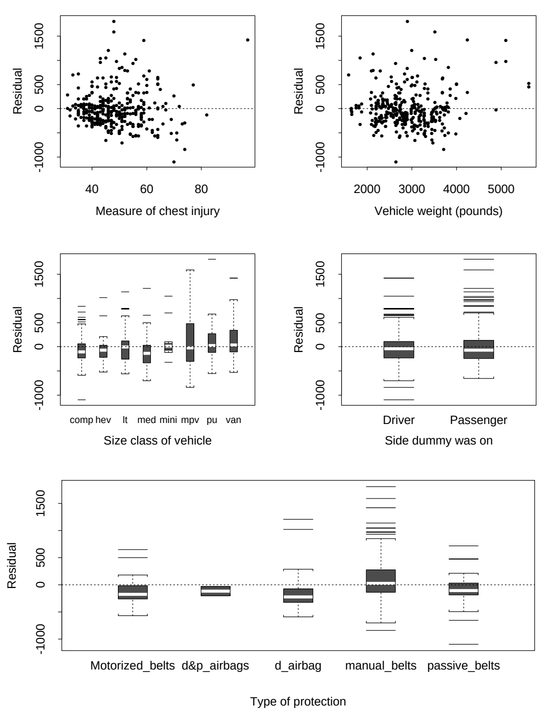
This also equals the square of the correlation between x and y — hence the name "r^2".

Crashtest Data: Regression of Head Injury on Chest Injury

```
The regression equation is
head = - 432 + 27.6 chest

334 cases used 18 cases contain missing values

Predictor        Coef       StDev          T        P
Constant       -431.7       108.7      -3.97    0.000
chest          27.583       2.199      12.55    0.000

S = 386.0      R-Sq = 32.2%     R-Sq(adj) = 32.0%
```



Residual Plots



Crashtest Data: Regression of Head Injury on Vehicle Weight

```
The regression equation is
head = 431 + 0.162 weight

340 cases used 12 cases contain missing values

Predictor        Coef       StDev          T        P
Constant        430.8       117.6       3.66    0.000
weight        0.16179     0.03939       4.11    0.000

S = 454.4      R-Sq = 4.8%      R-Sq(adj) = 4.5%
```