regressor x. Because $0 \le SS_{Res} \le SS_T$, it follows that $0 \le R^2 \le 1$. Values of R^2 that are close to 1 imply that most of the variability in y is explained by the regression model. For the regression model for the rocket propellant data in Example 2.1, we have

$$R^2 = \frac{SS_R}{SS_T} = \frac{1,527,334.95}{1,693,737.60} = 0.9018$$

that is, 90.18% of the variability in strength is accounted for by the regression model.

The statistic R^2 should be used with caution, since it is always possible to make R^2 large by adding enough terms to the model. For example, if there are no repeat points (more than one y value at the same x value), a polynomial of degree $n - 1$ will give a "perfect" fit ($R^2 = 1$) to n data points. When there are repeat points, R^2 can never be exactly equal to 1 because the model cannot explain the variability related to "pure" error.

Although R^2 cannot decrease if we add a regressor variable to the model, this does not necessarily mean the new model is superior to the old one. Unless the error sum of squares in the new model is reduced by an amount equal to the original error mean square, the new model will have a larger error mean square than the old one because of the loss of one degree of freedom for error. Thus, the new model will actually be worse than the old one.

The magnitude of R^2 also depends on the range of variability in the regressor variable. Generally R^2 will increase as the spread of the x's increases and decrease as the spread of the x's decreases provided the assumed model form is correct. Hahn [1973] observes that the expected value of R^2 from a straight-line regression is approximately

$$E(R^2) = \frac{\hat{\beta}_1^2 S_{xx}}{\hat{\beta}_1^2 S_{xx} + \sigma^2}$$

Clearly the expected value of R^2 will increase (decrease) as S_{xx} (a measure of the spread of the x's) increases (decreases). Thus, a large value of R^2 may result simply because x has been varied over an unrealistically large range. On the other hand, R^2 may be small because the range of x was too small to allow its relationship with y to be detected.

There are several other misconceptions about R^2. In general, R^2 does not measure the magnitude of the slope of the regression line. A large value of R^2 does not imply a steep slope. Furthermore, R^2 does not measure the appropriateness of the linear model, for R^2 will often be large even though y and x are nonlinearly related. For example, R^2 for the regression equation in Figure 2.3b will be relatively large even though the linear approximation is poor. Remember that although R^2 is large, this does not necessarily imply that the regression model will be an accurate predictor.