Supplemental Worksheet: Computing correlation coefficients

The formula for calculating R^2 given in *Should We Risk It?* only applies when you've fit the curve using least squares. When this hasn't been done, the symmetry TSS = ESS + RSS doesn't hold. Instead, use r, or the correlation coefficient.

To calculate correlation coefficient

$$r = \frac{\text{covariance (predicted, observed)}}{\text{standard deviation (predicted)} \times \text{standard deviation (observed)}}.$$

$$r = \frac{\mathrm{cov}(p, o)}{s_o \times s_p}$$

Recall that

$$s_p = \sqrt{1/(n-1)\sum (\bar{x}_p - x_{pi})^2}$$

$$s_o = \sqrt{1/(n-1)\sum (\bar{x}_o - x_{oi})^2}$$

and

$$\mathrm{Cov}(o,p) = 1/(n-1)\sum (\bar{x}_o - x_{oi})(\bar{x}_p - x_{pi})$$

The significance of r can be tested using the Student's t (n-2 degrees of freedom)

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Alternate formula for r:

$$r = \frac{n(\sum xy) - \sum x \sum y}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}}$$

(substitute x_{oi} for x and x_{pi} for y)