Formulas Relating to the Coefficient of Correlation of a Sample

n = sample size $\qquad \bar{x} = \frac{\sum x}{n} \qquad \bar{y} = \frac{\sum y}{n}$

$$s_x^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} \qquad s_x = \sqrt{s_x^2}$$

$$s_y^2 = \frac{\sum(y-\bar{y})^2}{n-1} = \frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1} \qquad s_y = \sqrt{s_y^2}$$

Covariance:

$$s_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$$

$$z_x = \frac{x-\bar{x}}{s_x} \qquad z_y = \frac{y-\bar{y}}{s_y}$$

Formulas for r (Pearson's Coefficient of Correlation)

$$r = \frac{\sum z_x z_y}{n-1}$$

$$r = \frac{s_{xy}}{s_x s_y}$$

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)\, s_x s_y}$$

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

Note: All of these formulas for r are equivalent. That is, they all produce the same value of r.

Additionally, if $(n-1)$ is replaced by n throughout all of the calculations, the value of r is unchanged.

The Coefficient Of Determination, R^2 is calculated as r^2. So, $R^2 = r^2$. (i.e., R and r are the same guy.)

Least Squares Line

yes, that is a "hat"

Equation of the line: $\hat{y} = a + bx$ where: $b = r\frac{s_y}{s_x}$ and $a = \bar{y} - b\bar{x}$

Residuals

note the "hat"

For each value of x, there is a residual error from the least squares line: $e = y - \hat{y}$.

The standard deviation of the residuals is: $s_e = \sqrt{\frac{\sum e^2}{n-2}} = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}}$

We define $(1 - r^2)$ to be the variation left in the residuals.

Note that:

$\sum y = \sum \hat{y}$, so

$\sum e = 0$, or $\bar{e} = 0$