Correlation and Regression Formula Review

I. Correlation formulas

A. Coefficient of correlation $$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n(\sum X^2) - (\sum X)^2][n(\sum Y^2) - (\sum Y)^2]}}$$

B. Coefficient of determination $r^2 = (r)^2$

C. Coefficient of nondetermination $\bar{r}^2 = 1 - r^2$

D. The value of t when determining the significance of the coefficient of correlation r $$t = \frac{r - \rho}{\sqrt{\frac{1-r^2}{n-2}}}$$ and df = n - 2

II. Regression formulas

A. The regression equation $\hat{y}_{.x} = a + bx$

B. The slope of the regression equation $$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

C. The y-intercept of the regression equation $$a = \bar{Y} - b\bar{x} = \frac{\sum y}{n} - b\frac{\sum x}{n}$$

D. The standard error of the estimate $$S_{y.x} = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n-2}} = \sqrt{\frac{\sum Y^2 - a(\sum Y) - b(\sum XY)}{n-2}}$$

E. An interval estimate for the conditional mean of y for some given value for x

$$\hat{y}_{.x} \pm ts_{y.x} \quad \text{or} \quad \hat{y}_{.x} \pm ts_{y.x}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

Note: An interval estimate for an individual value of y, sales for a recently hired 24-year-old salesperson or grades for your roommate who studied 5 hours, would require adding a 1 under the radical. This makes the interval substantially larger.

$$\hat{y}_{.x} \pm ts_{y.x}\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$