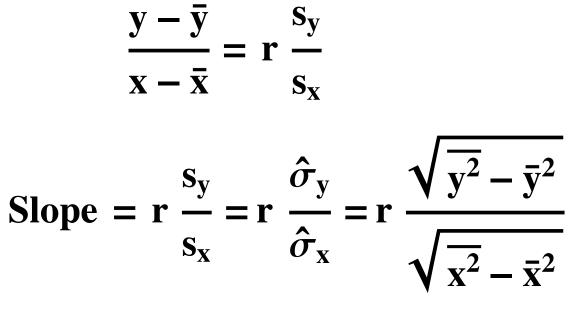
Lecture outline 2 - 13 - 09

Part of the period will cover numerical examples as in 2-11-09. The rest will be devoted to the points below.

1. Important characterization of all points (x, y) which lie on the line of regression :



pg. 197

2. Taking x = 0 in
$$\frac{y - \bar{y}}{0 - \bar{x}} = r \frac{s_y}{s_x}$$

gives

intercept =
$$\bar{y} - \bar{x}$$
 slope

pg.198

3. For every x, solving for y in $\frac{y - \bar{y}}{x - \bar{x}} = r \frac{s_y}{s_x}$ gives predicted y = pt on regr line : pred y = $\bar{y} + (x - \bar{x})$ slope

4. For an approximately ELLIPTICAL plot, at a given x the distribution of y is approximately NORMAL with

mean = predicted y

std dev = $\sqrt{1 - r^2} S_y$

Notice that the mean depends upon x but the std dev does not.

5. For an ELLIPTICAL PLOT the regression predictor $\bar{y} + (x - \bar{x})$ slope is optimal in the sense of least mean squared error of prediction.

5. r^2 is exactly the fraction of $\hat{\sigma}_y^2$ explained by the sample regression.

pp. 204-06 (Note text also uses R for r. The Greek "rho" ρ is also used for r).