Section 3.2

Linear Regression

PART IV

→ Develop the slope & y-int. of the LSR line.

→ Based on 2 assumptions:

1. The centroid: $(\bar{x}, \bar{y}) = \left(\frac{\Sigma x}{n}, \frac{\Sigma y}{n}\right)$ must be on the line.
2. S_y = std. dev. of all y values & S_x = std. dev. of all x values should be used to help determine the rise & run of the slope of the LSR line.

→ If the data points exhibit perfect linear correlation (+ or −) then the LSR line's slope should be $\pm \frac{\text{Rise}}{\text{Run}} = \pm \frac{S_y}{S_x}$. (when $r = \pm 1$).

→ When the points don't exhibit perfect linear correlation, we want the slope to have a diminishing influence. In fact as r becomes close to zero, we want the slope to approach zero. $\underbrace{\hat{y} = a + bx \approx a}_{\text{slope } b \to 0}$ ($b \approx 0$)

→ x influence vanishes.

→ Slope $b = r \cdot \frac{S_y}{S_x}$ ✓

→ Point slope form: $y - y_0 = m(x - x_0)$

LSR Line

$\hat{y} = a + bx$

slope: $b = r \cdot \frac{S_y}{S_x}$

$y - \bar{y} = b(x - \bar{x})$

$y = bx + \underbrace{\bar{y} - b\bar{x}}_{\text{y-intercept "a"}}$

y-int: $a = \underbrace{\bar{y} - b\bar{x}}$.

note that as the slope approaches zero (because r is approaching zero), the y-intercept approches $\bar{y}$.