this time: correlation & next time: regression

read: ch. 6 in lecture notes pp. L-214 – L-268

AMS 7 26 Feb 09

①

lab 5 due by 5pm Fri 27 Feb

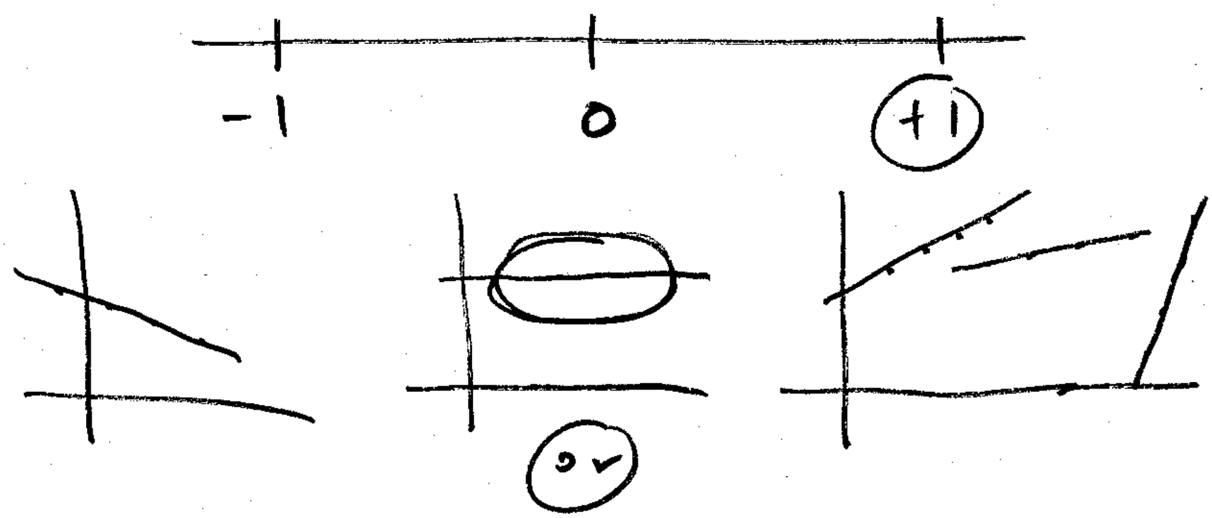
hwk 4 due Thu 5 Mar

correction: lab 6 due by 5pm Fri 6 Mar

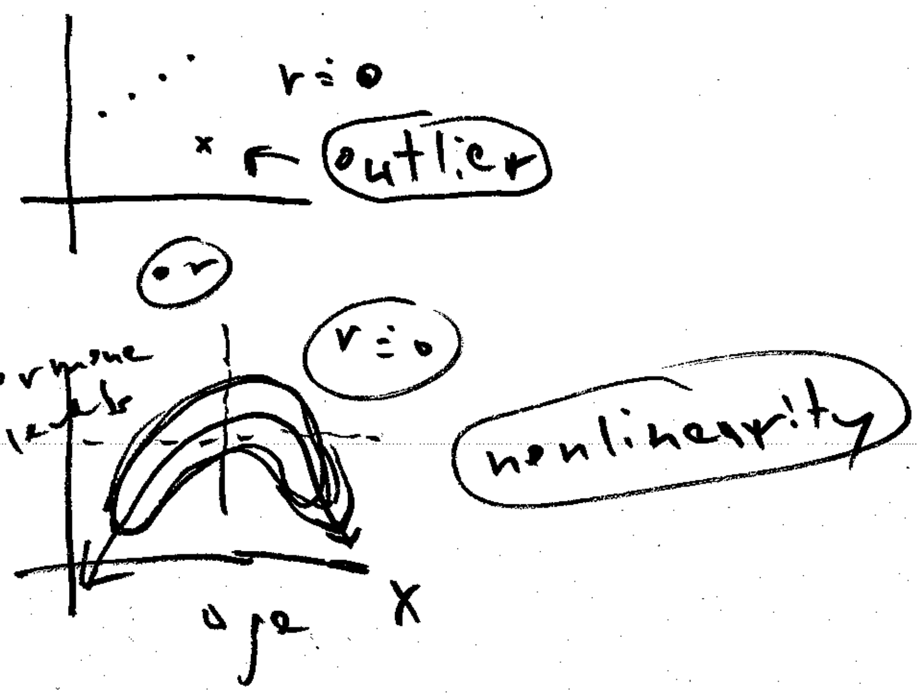
today: L-221 → L-2

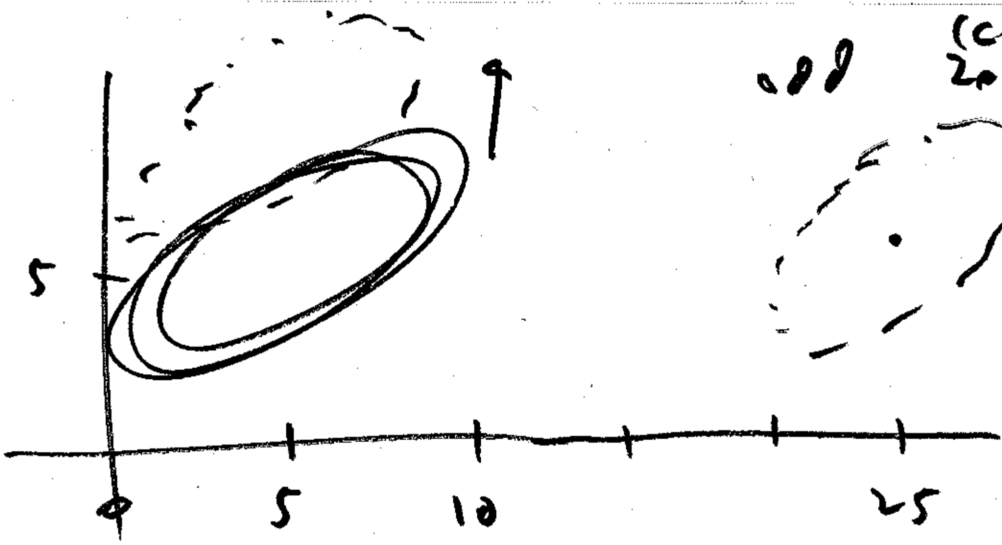
facts about r

①

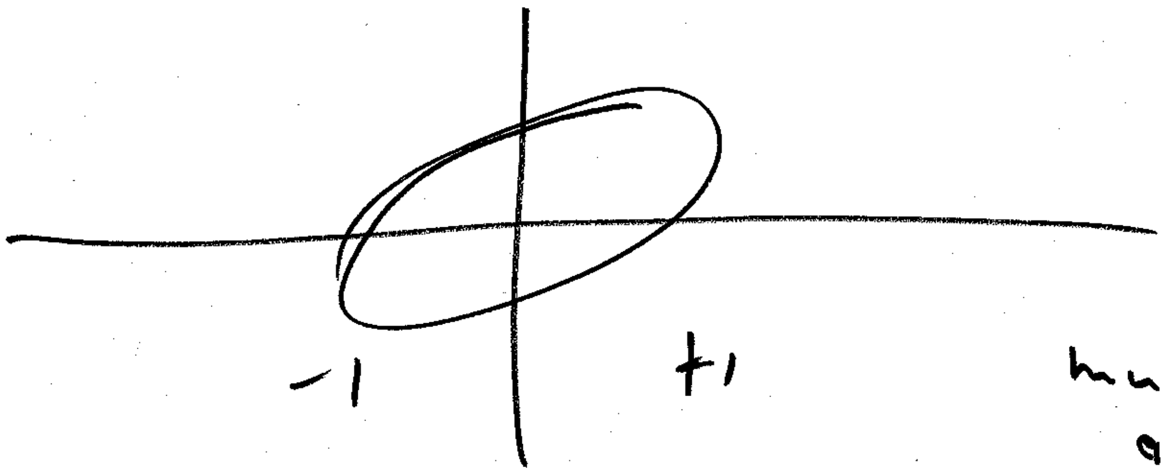


$$r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x^*} \right) \left(\frac{y_i - \bar{y}}{s_y^*} \right)$$

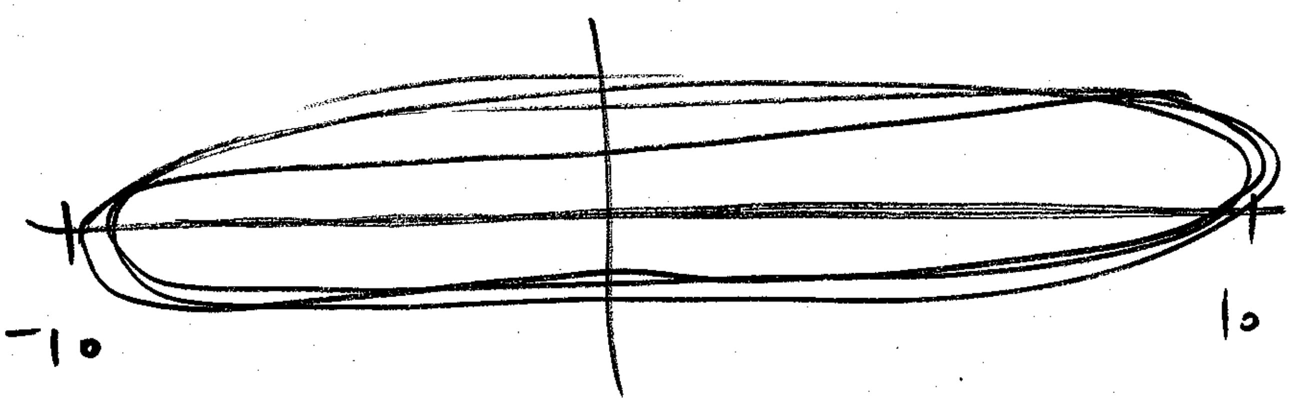


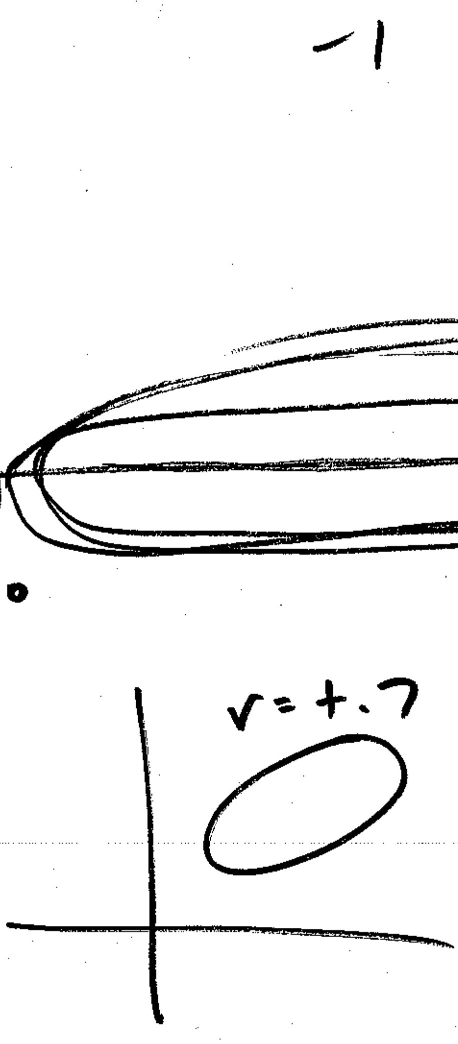


add (c) 20 to all x-values leaves r unchanged

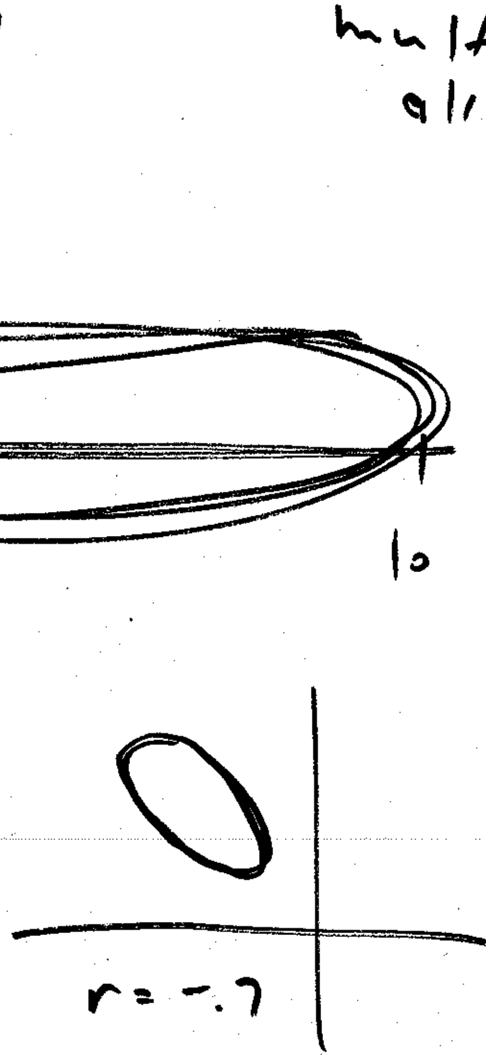


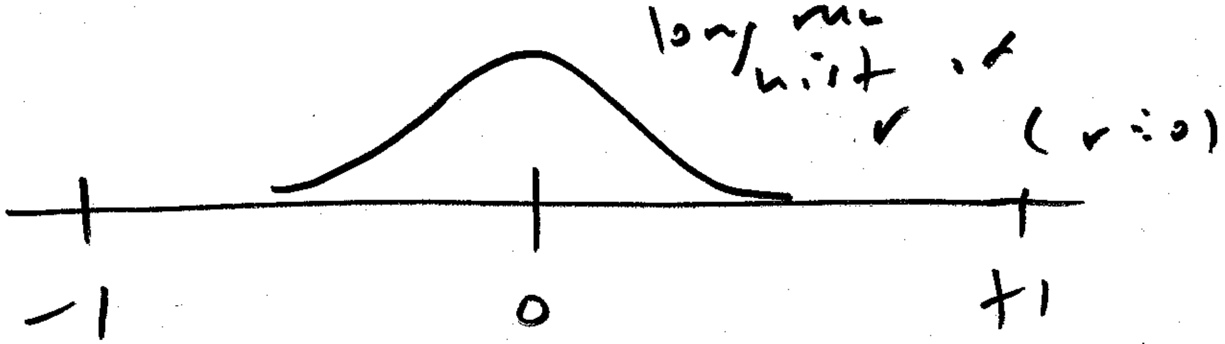
multiply all x values by 10

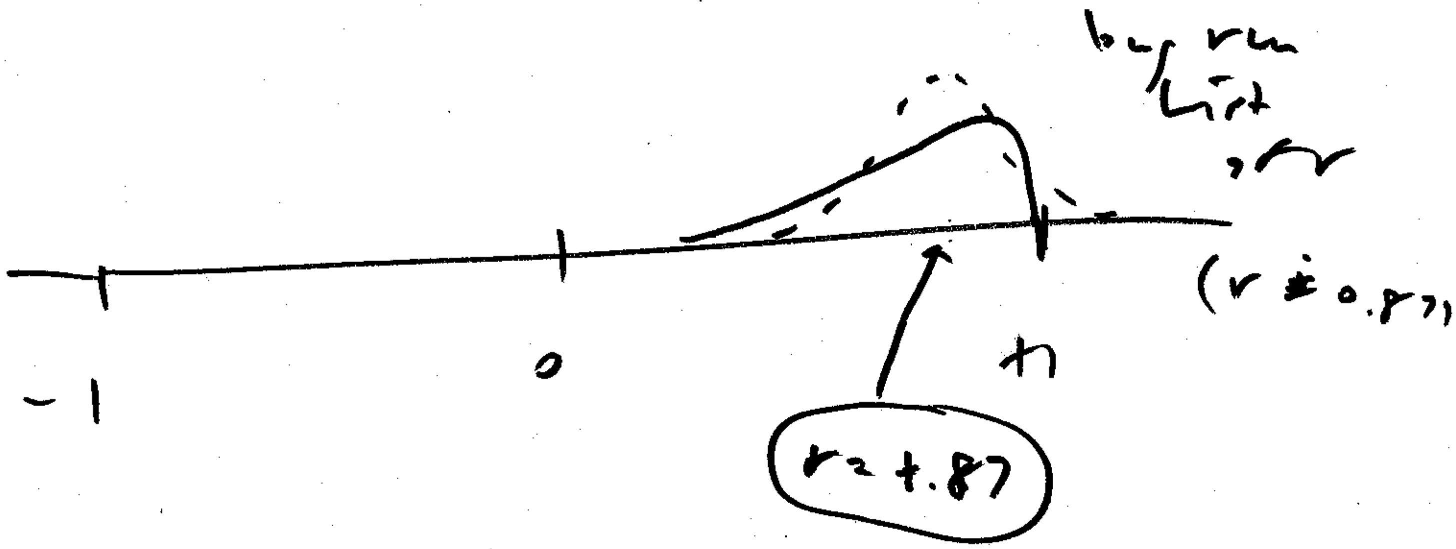




mult. x by -1







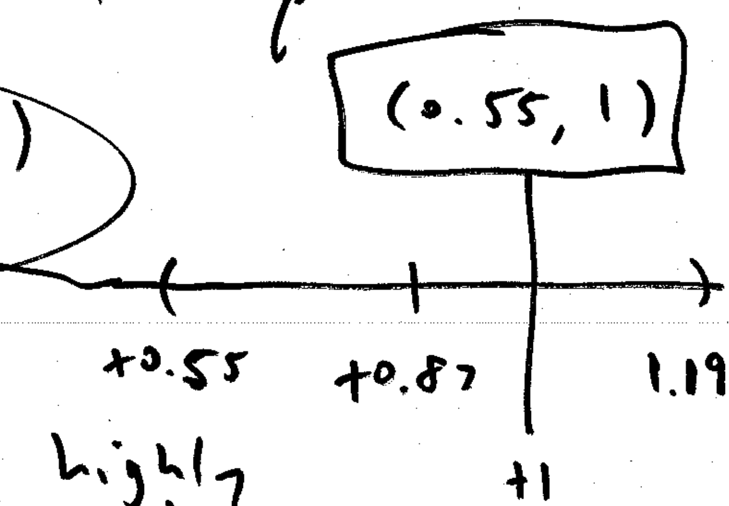
skip: L-231 → L-244

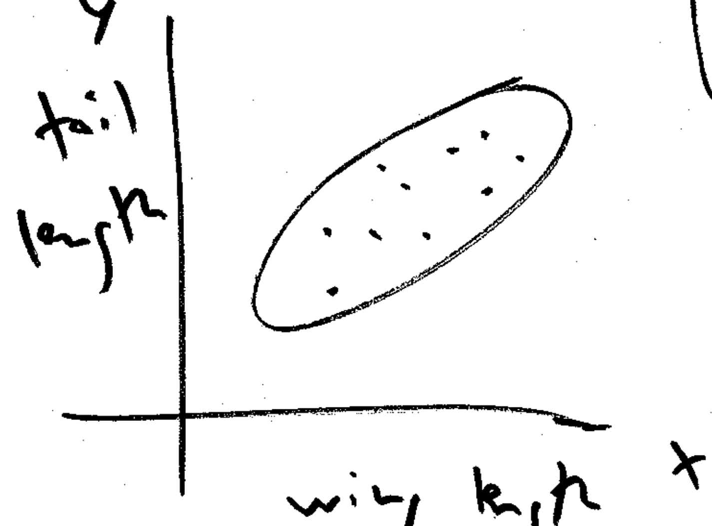
approx 98% CI for ρ:

$r \pm 1.96 \, \widehat{SE}(r)$

$(+0.87) \pm 1.96 \, (0.16)$

± 0.32

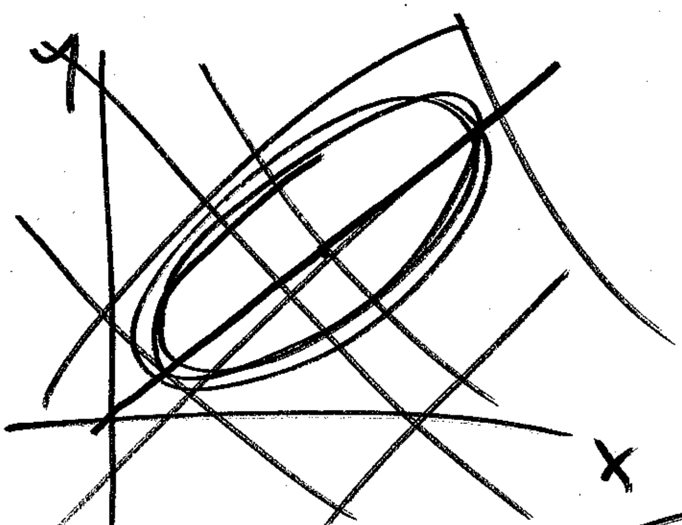




$r = +0.87$

Q1: how use x to predict y?

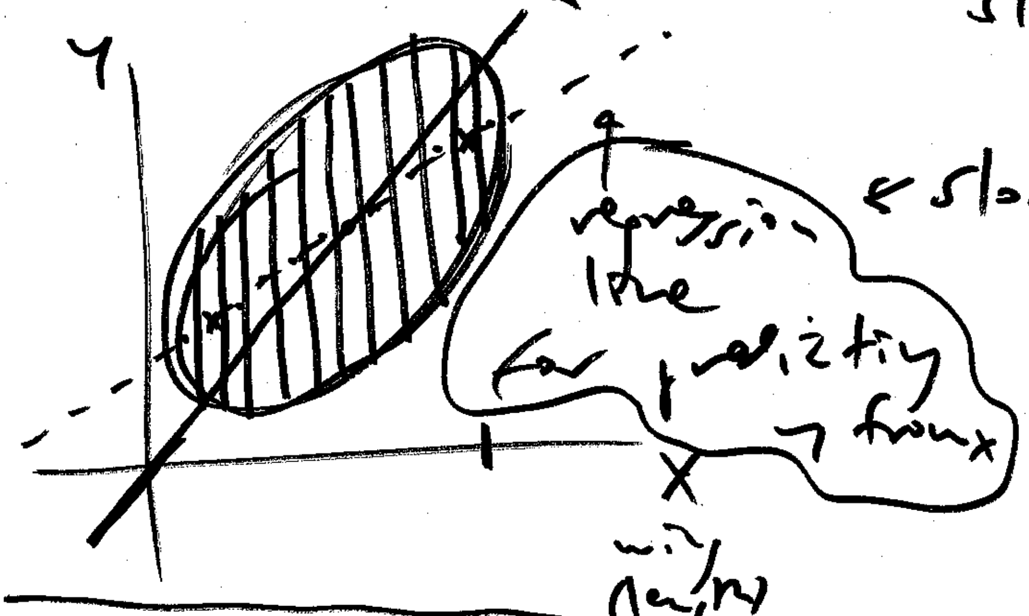
Q2: what's the equation of best line for predicting y from x?



(major axis)

Galton (1890s)

SD line: slope $\frac{s_y}{s_x}$



← slope $r \cdot \frac{s_y}{s_x}$

y-intercept

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

equation of best line for predicting y from x

intercept! slope $\hat{\beta}_1 = r \cdot \frac{s_y}{s_x}$ ① ; line

goes through $(\bar{x}, \bar{y})$:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$ so

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$ ②