AMS 7
12 Nov 15

this time: correlation & regression

next time:

read: LN pp. L-245 – L-268

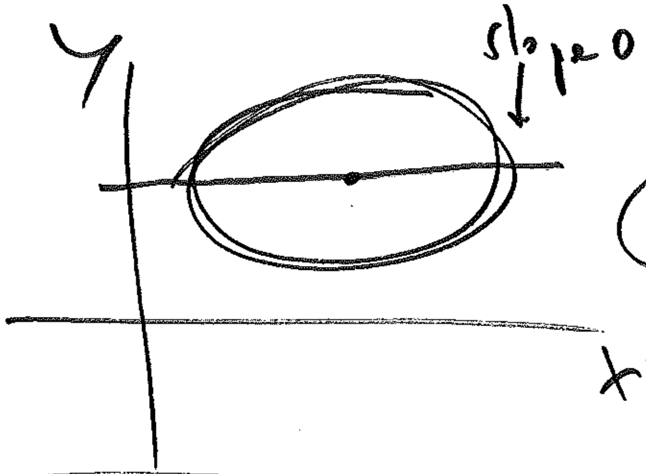
1

this time: LN pp. L-225 →

hwk 4 due next Thu (19 Nov) in class

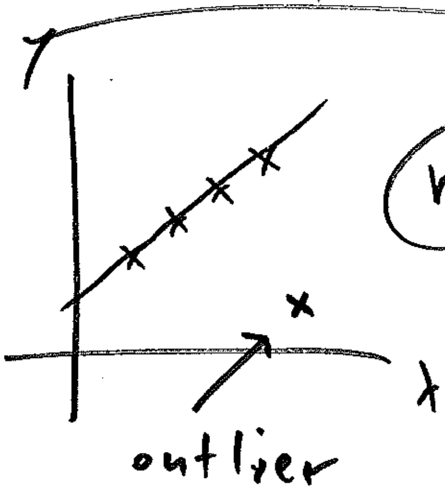
facts about r

$r = 0$ → 3 different scatterplot shapes



$r \doteq 0$

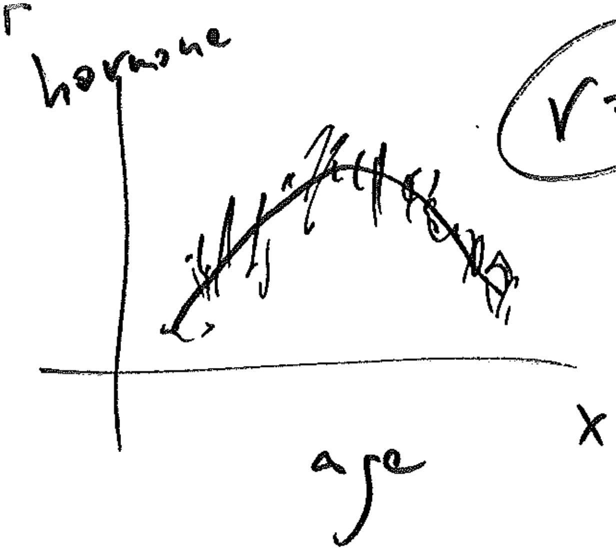
no linear assoc. → x not useful in predicting y



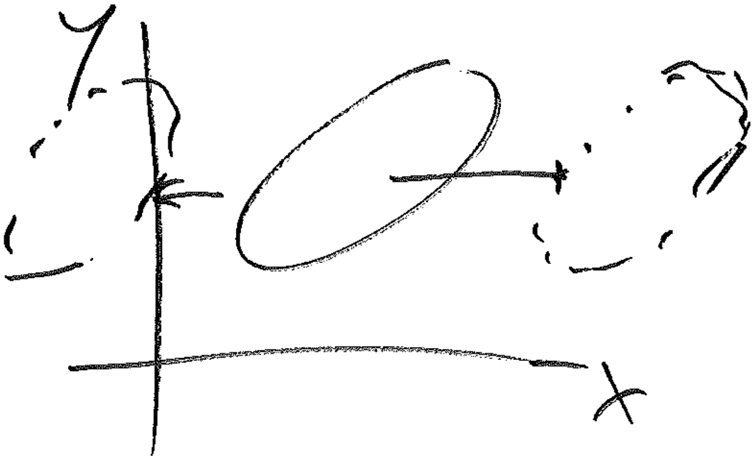
$r \doteq 0$

outliers can distort r, especially with small n

②

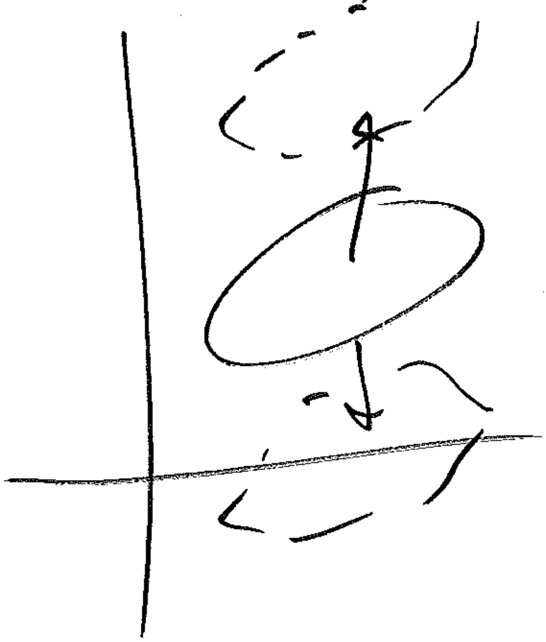


$r=0$ strong non-linear relationship (quadratic)



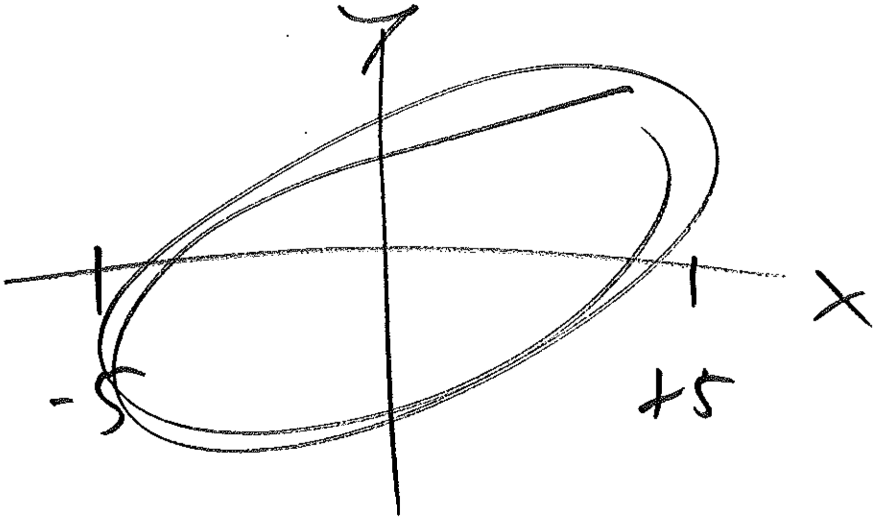
add a constant ($c>0$) to x
($c<0$)

leaves r unchanged

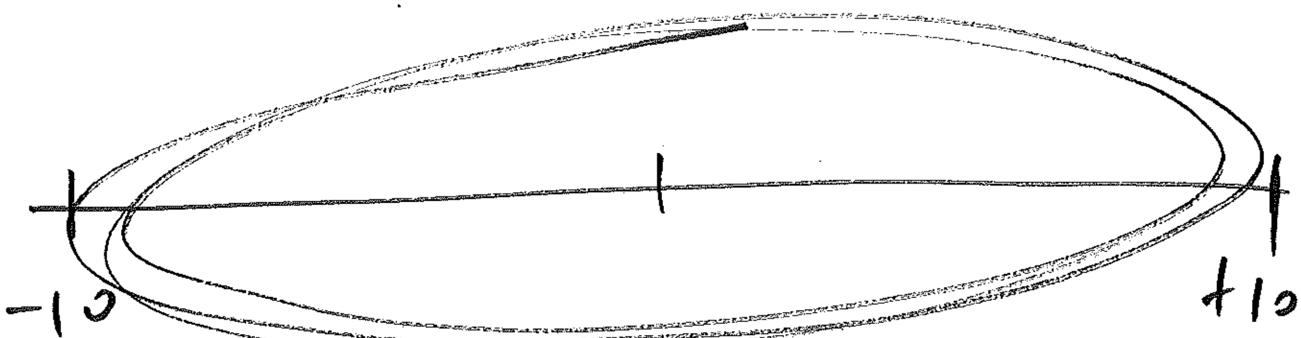


add a constant ($c>0$)
($c<0$)
to y

leaves r unchanged



mult. diff x by 2



$r = +.87$ (wing l., tail l.) sparrows

$\boxed{Q:}$ Is this r large in practical terms?

$\boxed{A:}$ $x = 10 \rightarrow \hat{y} = 7$ cm

smallest

predicted

tail l.

$x = 11.5 \rightarrow \hat{y} = 8.25$ cm

largest

so r is practical

this diff. is large in practical terms

$r = +.87$

$\hat{SE}(r) = 0.16$

$\pm .32$

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

large-n approx. 95% CI

95% CI for ρ

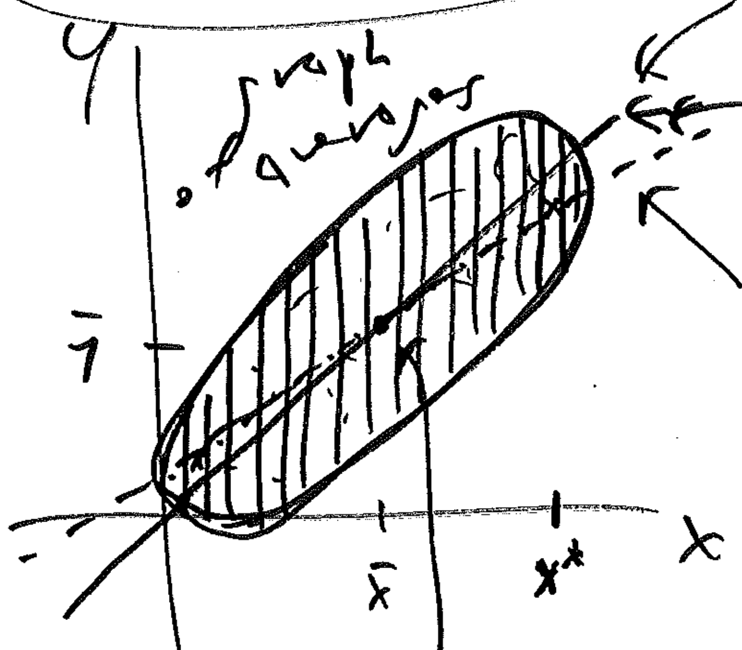
0 .55 .87 1.19

$(0.55, +1)$

truncate at 1.00

predicting y from x

Galton



SD line

slope $\frac{s_y}{s_x}$

regression line:

slope $r \cdot \frac{s_y}{s_x}$

best line for predicting y from x

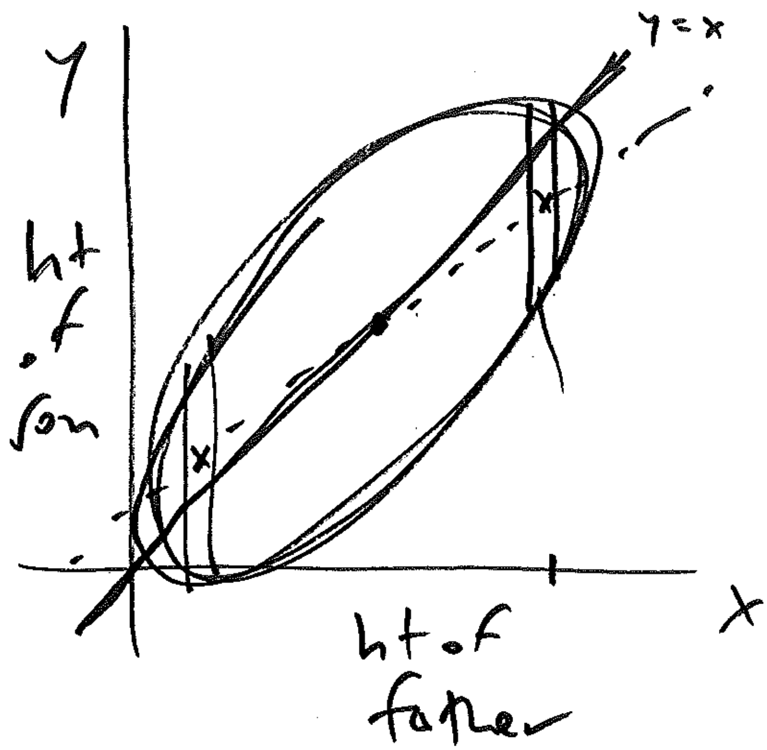
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

← regression line equation

y-intercept ($\hat{\beta}_0$), slope ($\hat{\beta}_1$)

$$\hat{\beta}_1 = r\,\frac{s_y}{s_x}$$

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

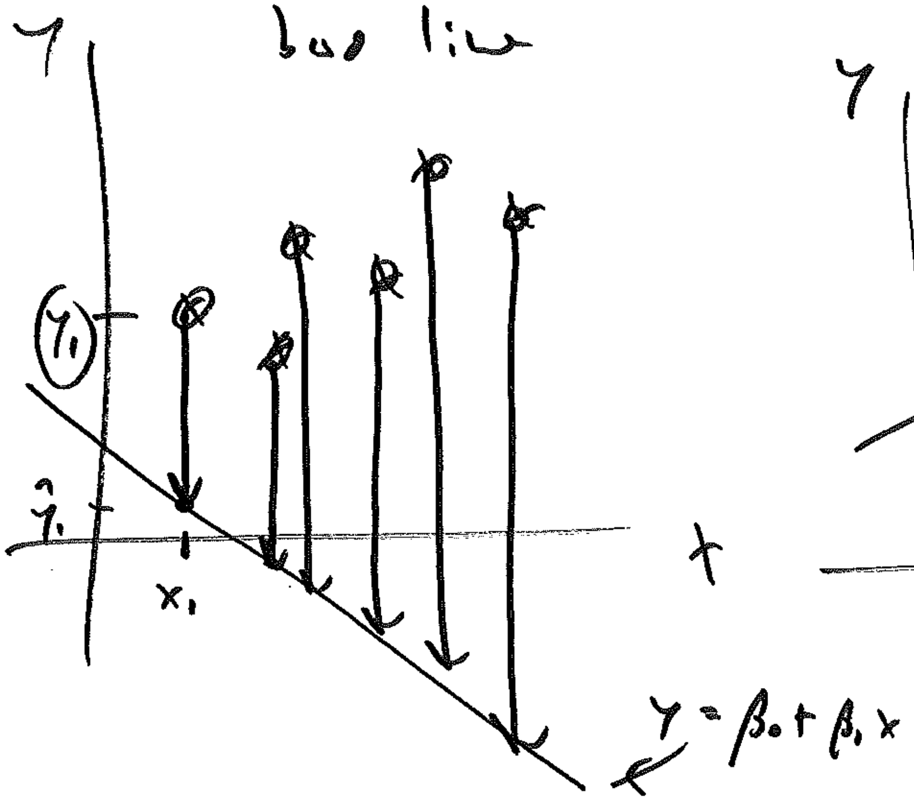


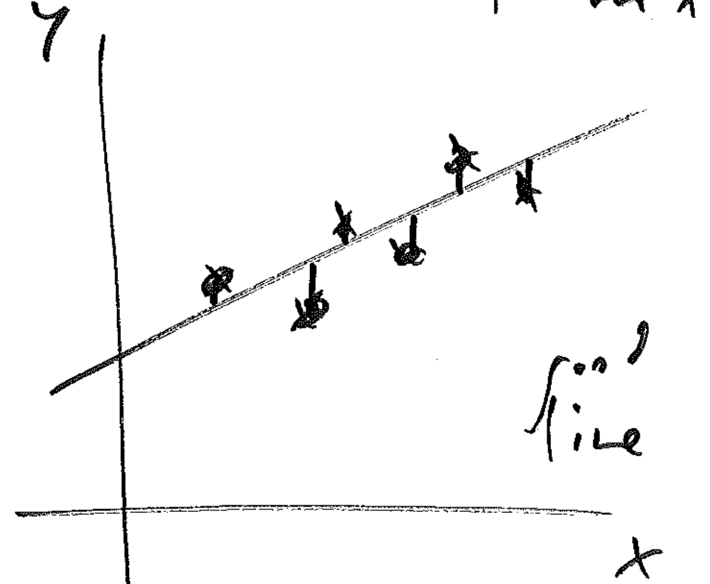
$n \doteq 1000$ families w/ at least 1 son

∘ random son → ht

father → ht

tall fathers have tall sons, but not as tall another way to get the best line for pred. Y from x





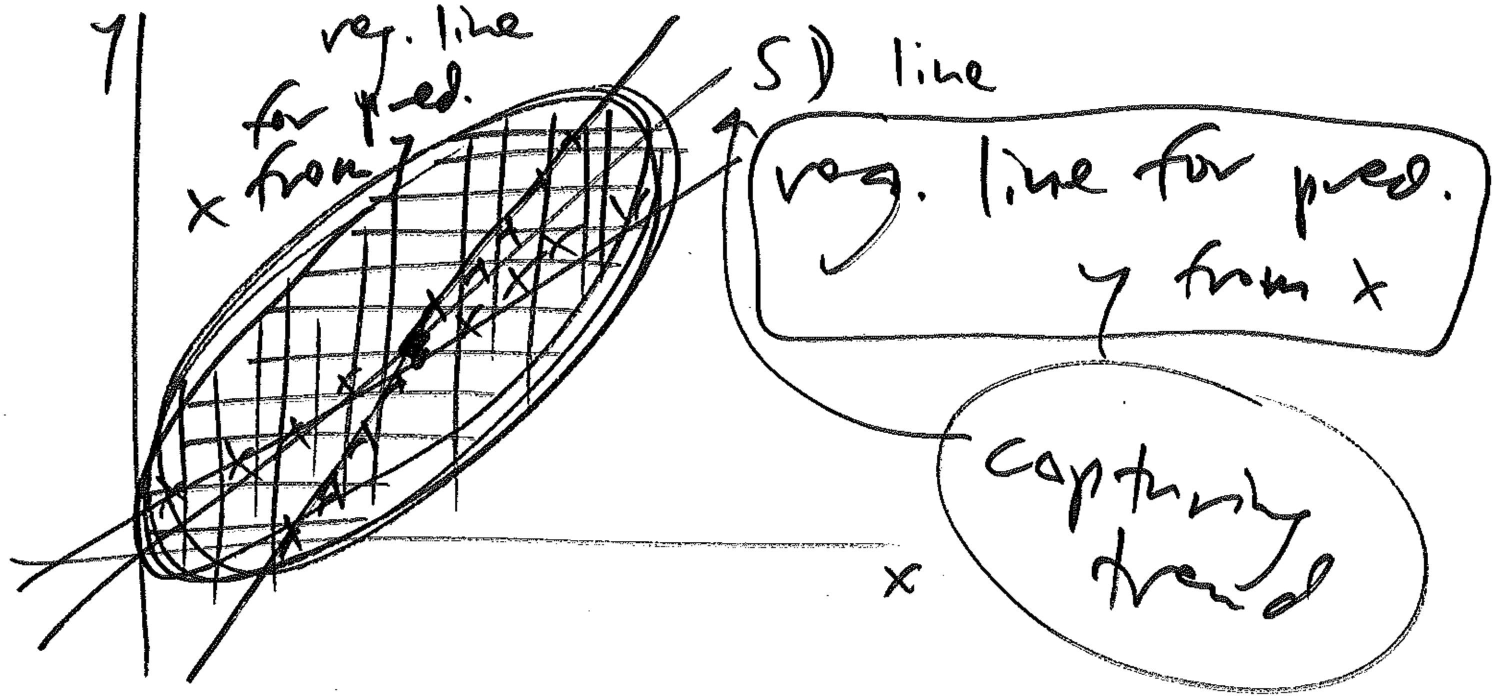
$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - (\beta_0+\beta_1 x_i)\right]^2$$

find β_0, β_1 to minimize

result: least squares line

(Gauss) 1800

math fact: regression line = least squares line


y
reg. line for pred. x from y
SD line
reg. line for pred. y from x
capturing trend
x