AMS 7
24 May 2018

this time: correlation & regression

next time: ANOVA

read: LN pp. L-214 + L-221

today: LN pp. L-214 + L-224

hwk 3 due by 11.59 pm on Sun 27 May 18 at canvas

(1)

$r(x,y) = r(y,x)$ ✓

$r(x,x) = +1$

inference with r

$0.87 \pm 2(0.08) = (0.71, 1.0)$ (0.16)

approx 95% CI for ρ

0 — 0.71 — 0.87 — 1.03 — 1.0

null

statistic ✓

L-245

regression

the regression effect

regression to the mean

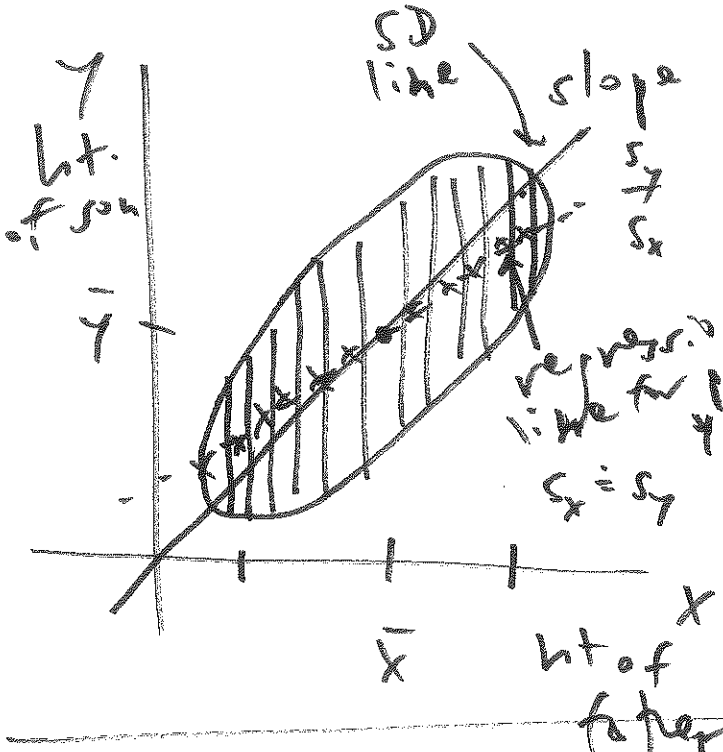
+

2nd

$\bar{Y}$

1st

$\bar{X}$

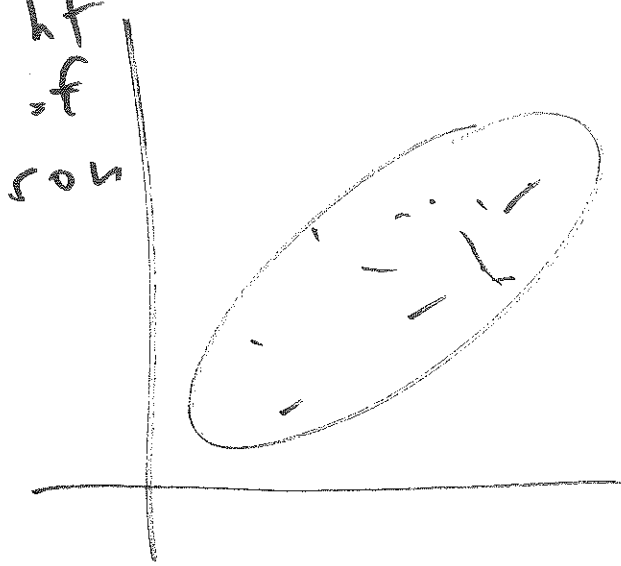


$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{y}$: predicted y value; $\hat{\beta}_0$: est. intercept; $\hat{\beta}_1$: est. slope

(1890 Galton)

(1)



UK
n = 1000 families with at least one son

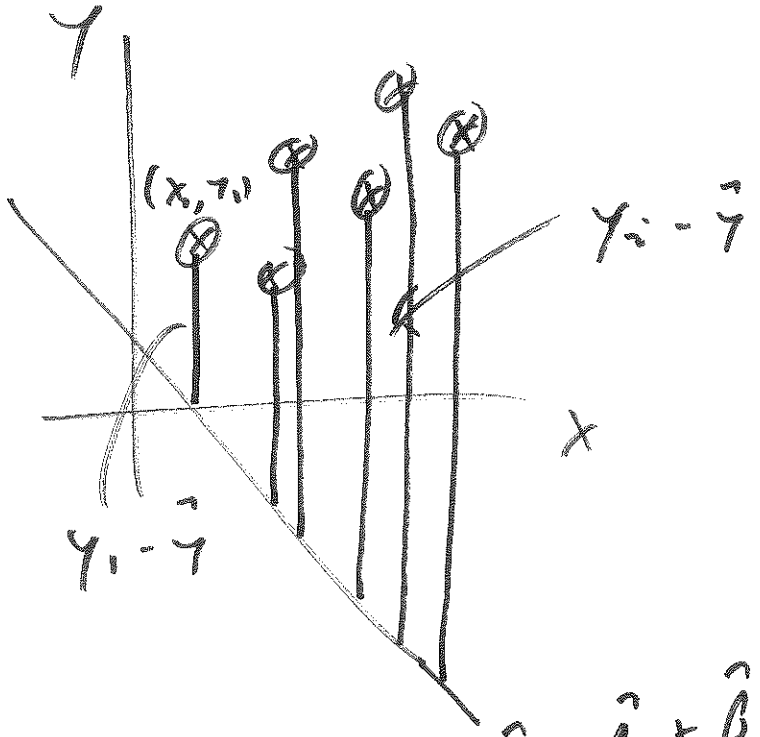
regression to mediocrity

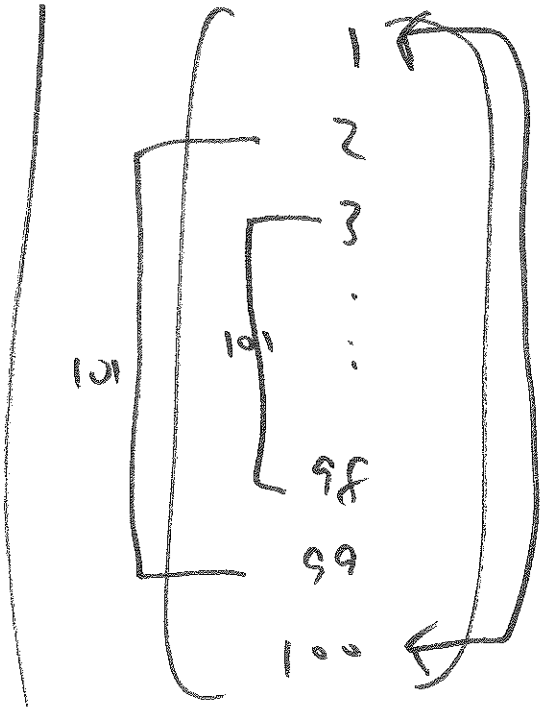
$$\hat{\beta}_1 = r \cdot \frac{s_y}{s_x}$$

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

(1800)

Gauss





$(50)(101) = 5050$

find $(\hat{\beta}_0, \hat{\beta}_1)$ to minimize

$$\sum_{i=1}^{n} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

$(\hat{\beta}_0, \hat{\beta}_1)$: least squares line

regression line = best fit line