AMS 7 - Lecture 5/24/18

This Time: Correlation & regression
Next Time: ANOVA

HWK 3 due by 11:59pm on Sun 27 May on canvas

→ Read: LN Pg L-214 → 282
Today: LN Pg L-214 →

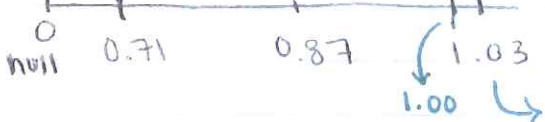
L-221 ▷ (Simple) Correlation (Simple Linear) & Regression

Math Fact
↳ Interchanging the role of x & y does not change the correlation
$r(x,y) = r(y,x)$ $r(x,x) = +1$

- Inference w/ r
Population correlation # (ρ) "RHO"
Sample correlation # (r)

L-230 $0.87 \pm \underbrace{2(0.08)}_{0.16}$

approx 95% CI for ρ

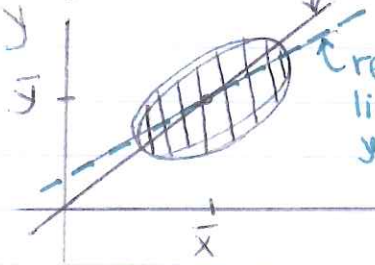


Chop off @ 1.00

Interval runs from (0.71, 1.0) to

Devil's adv. → null story statsig ✓

L-239 & R-25: extra credit for HW #4
* Skip to pg. 245 in organized Lecture notes
▷ Regression SD line (slope s_y/s_x)



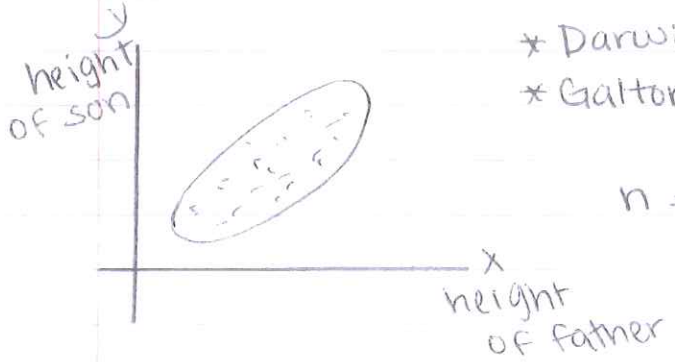
L-245

* Figure out the slope of the best line

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

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

5/24/18



* Darwin + Mendel ⇒ Fisher
* Galton (1890)

n = 1000 families with atleast 1 son

L-246

• regression "to mediocrity"

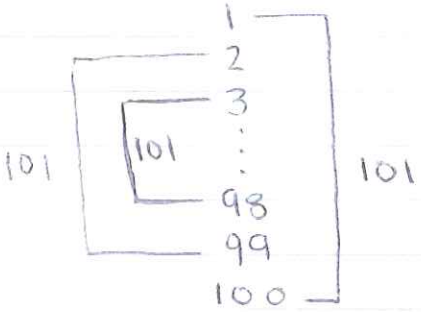
$$\hat{\beta}_1 = r \cdot \frac{s_y}{s_x}$$

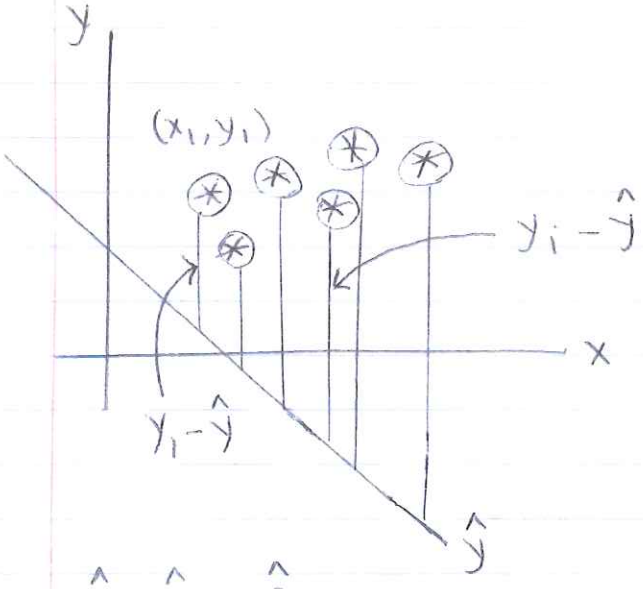
R-25 ← Formula Sheet

▷ Gauss (1800)

Adding #s from 1-100

(50)(101) = 5050

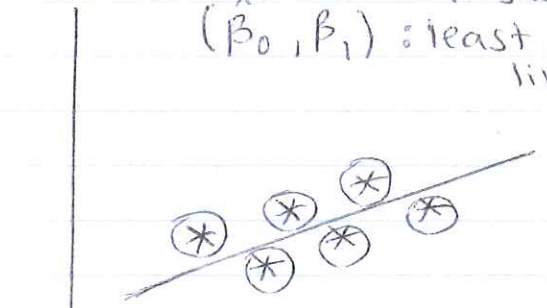




$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Not the best line

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



Better line

$$\sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x))$$

→ abs. value

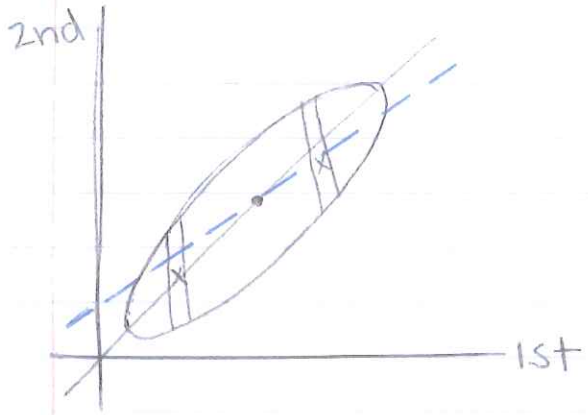
*Find $(\hat{\beta}_0, \hat{\beta}_1)$ to minimize,

5/24/18

[3]

MATH FACT: regression line = least squares line

GRE test scores ex



* first GRE score vs. Second GRE score

*regression to the mean

→ those w/ higher scores tend to get a worst grade the 2nd try & those with a low score tend to do better the 2nd time taking the GRE