STT 2-4-09a. Key to one variant of EXAM 1 has been posted. Discuss Friday.

PERIOD OF TIME 2 RECITATIONS FROM NOW.

Chapters 7, 8, 9.

Up to now we've looked at 1 variable x.

Ch 7–9 concerned w/ more than one x, y.

QUES: $\left[\begin{array}{l} x = \text{HS GPA} \\ y = \text{Classification of HS} \end{array}\right]$ w = 1ST YR GPA MSU.

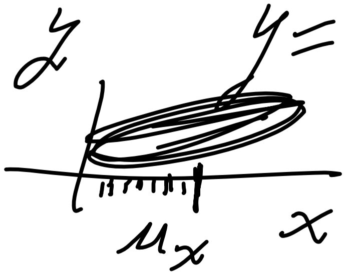
or

x = 2008 PROPERTY TAX

y = 2009 " TAX

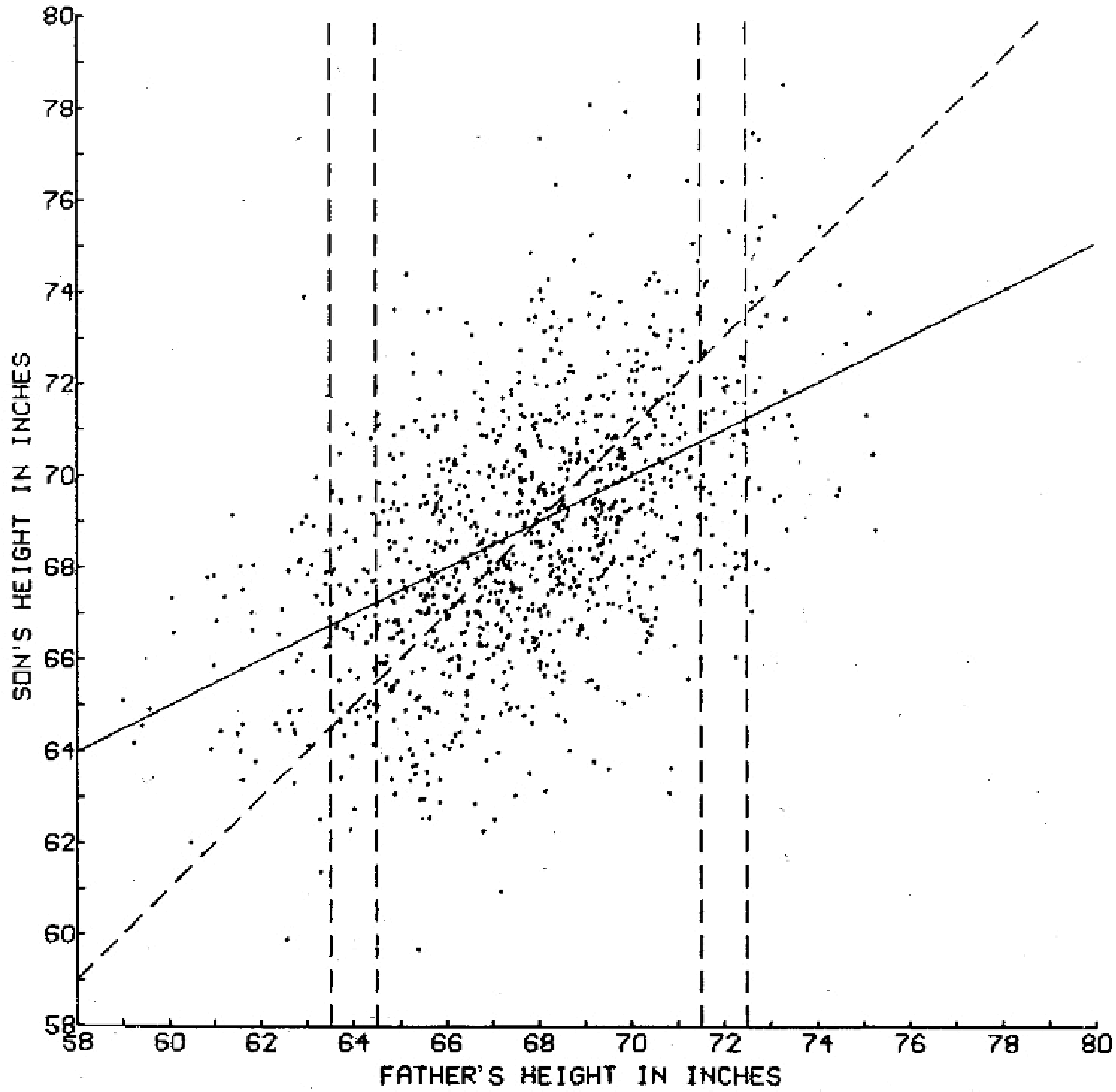
SAMPLES OF PROPERTIES

x_i, y_i — AUDIT DETERMINED

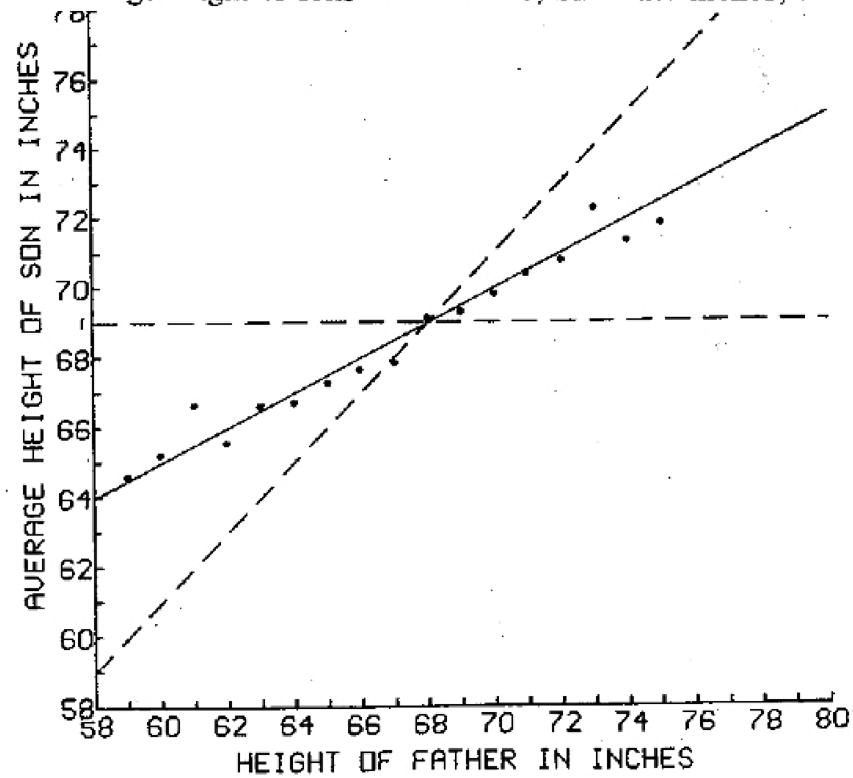


OBJECTIVE μ_y $\bar{y}$??

SUGGESTS $\bar{y}$ IS LOW. — FIND A METHOD OF "BOOSTING" $\bar{y}$


SON'S HEIGHT IN INCHES
80
78
76
74
72
70
68
66
64
62
60
58
58
60
62
64
66
68
70
72
74
76
78
80
FATHER'S HEIGHT IN INCHES

average height of fathers ≈ 68 inches, SD ≈ 2.7 inches
average height of sons ≈ 69 inches, SD ≈ 2.7 inches, $r \approx 0.5$



Freedman Pisani Purves, 1980

TWO POINTS;

(1) MAY WISH TO PROJECT y SCORE FOR AN INDIVIDUAL USING THEIR x SCORE eg y = MSU GPA, x = HS GPA.

or (2) WE'RE AFTER μ_y + HOPE TO USE (x_i, y_i) SCORES TO IMPROVE ON $\bar{y}$.

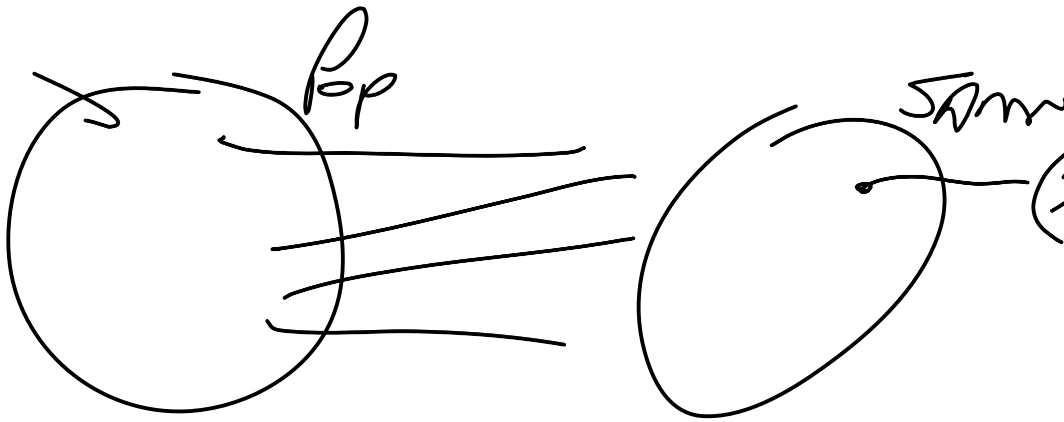
ANS TO (2) TWO COMPONENTS

(a) IMPROVED ESTIMATOR (IMPROVING UPON $\bar{y}$)

(b) EMOE + 95% CI FOR μ_y APPLICABLE TO THE IMPROVED ESTIMATOR.

TO DO THIS I NEED SAMPLE (x_i, y_i) $i \leq n$.

✱ I NEED TO KNOW μ_x.



(x_i, y_i) $\quad i \le n$

$\bar{x} \quad s_x$

$\bar{y} \quad s_y$

r (L.C.R)

CORRELATION

each $i \le n$

$$z_{x_i} = \frac{x_i - \bar{x}}{s_x}$$ STD SCORE OF INDIVIDUAL i ON X.

$$z_{y_i} = \frac{y_i - \bar{y}}{s_y}$$

$$r = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$$

IMPROVEMENT ON $\bar{y}$?

SOLVE $\dfrac{y - \bar{y}}{x - \bar{x}} = r \dfrac{s_y}{s_x}$

μ_x KNOWN (*)

GET $y = \bar{y} + \underbrace{(\mu_x - \bar{x})}_{+ \text{ IF } \mu_x > \bar{x}} r \dfrac{s_y}{s_x}$

COULD USE JUST $\bar{y}$

IF r IS + ALSO

IF $\mu_x > \bar{x}$
AND $r > 0$
THIS INCREASES $\bar{y}$

SO THE "IMPROVED ESTIMATOR OF μ_y"
IS JUST $\bar{y} + (\mu_x - \bar{x})\, r\, s_y/s_x$

SECONDLY, THE EMOE FOR "IMPROVED ESTIMATOR" IS

$$1.96\,\frac{s_y}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\sqrt{1-r^2}$$

RECALL EMOE $\bar{y}$ IS

$$1.96\,\frac{s_y}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

+ GUESS WHAT?

$$|r| \leq 1$$

EACH OF n SAMPLE INDIVIDUALS HAS x SCORE AND y SCORE

ASSOCIATED 95% CI FOR μ_y IS

$$\Big(\underbrace{\bar{y} + (\mu_x - \bar{x})\,r\,\frac{s_y}{s_x}}_{\text{"IMPROVED ESTIMATOR"}}\Big) \pm 1.96\,\frac{s_y}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\sqrt{1-r^2}$$

GALTON'S DATA CH7 - SAMPLE

ELLIPTICAL

SAMPLE PAIR $i = 10$

x_i y_i

FATHER'S HT

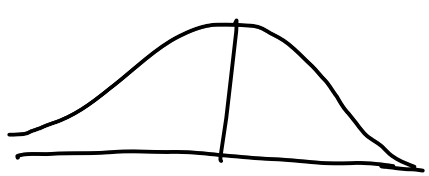
SON'S HT.

i^{TH} FATHER-SON PAIR $i \leq n$. PAIRS

x_i = HT OF FATHER

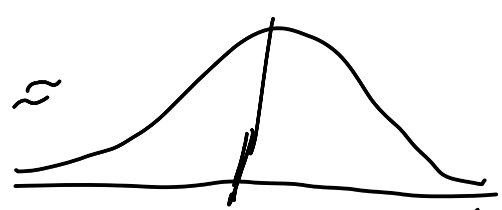
y_i = HT OF SON

$\approx$



$\approx 69''$ x FATHERS HTS

$\approx$



69.5 y SONS.

STT200 2-4-09b.

TODAY: BEGIN CH 7-9 REGRESSION CORRELATION

$\underbrace{\text{CH 7-9}}$ THROUGH 2 RECITATIONS.

SO FAR WE'VE LOOKED ONLY AT SAMPLE OF n FROM POP^N OF N. SINGLE SCORE X.

DATA CONSISTS OF $X_1, \ldots, X_n$

SCORE X OF THE FIRST SAMPLE INDIVIDUAL

X SCORE OF n-TH SAMPLE INDIVIDUAL

HAD EMOE FOR $\bar{X}$ (EST OF μ_X)

IS $1.96 \frac{\sigma_X}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

ASSOCIATED CI (95%): $\bar{X} \pm 1.96 \frac{\sigma_X}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

WITH CH 7–9 WE MOVE TO MORE THAN ONE VARIABLE.

TWO MAIN TASKS:

(1) Say x = HS. AVG y = MSU AVG.

PERHAPS WE NEED TO UNDERSTAND THE RELATIONSHIP – IN PARTICULAR WE MAY WISH TO PREDICT A GIVEN APPLICANT'S y FROM THEIR x.

(2) MAY WISH TO ESTIMATE μ_y = POP^N AVG SALES (PER SHOPPER) 2008 HOLIDAY SEASON. LET x = PERSON'S 2007 HOLIDAY PURCHASES. ALREADY KNOW μ_x. (*) LEADS TO IMPROVED EST OF μ_y. (IF $\bar{x} \ll \mu_x \Rightarrow$ SHOULD INCREASE $\bar{y}$)

I'LL LOOK AT (2) JUST TO GIVE YOU AN IDEA OF HOW IT GOES.

SAMPLE n INDIVIDUALS, RECORD x_i y_i

x_i: X-SCORE FOR iTH SAMPLE PERSON

y_i: Y SCORE FOR iTH SAMPLE PERSON

ALREADY KNOW SOMETHING ABOUT

$\bar{x}$ $\quad s_x$

$\bar{y}$ $\quad s_y$ $\quad r$

lowercase R - CORRELATION - $-1 \leq r \leq 1$

ENHANCED ESTIMATOR OF μ_y: $\bar{y} + (\mu_x - \bar{x}) \, r \, s_y / s_x$

$$\boxed{\text{EST MOE OF "ENHANCED"} \quad 1.96 \frac{s_y}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \sqrt{1-r^2}}$$

IF $\mu_x > \bar{x}$ KNOWN!

IF $r > 0$ POSITIVE ASSOC x, y.

SENSIBLE THAT $\bar{y}$ IN INCREASED.

SO 95% CI FOR μ_y IS

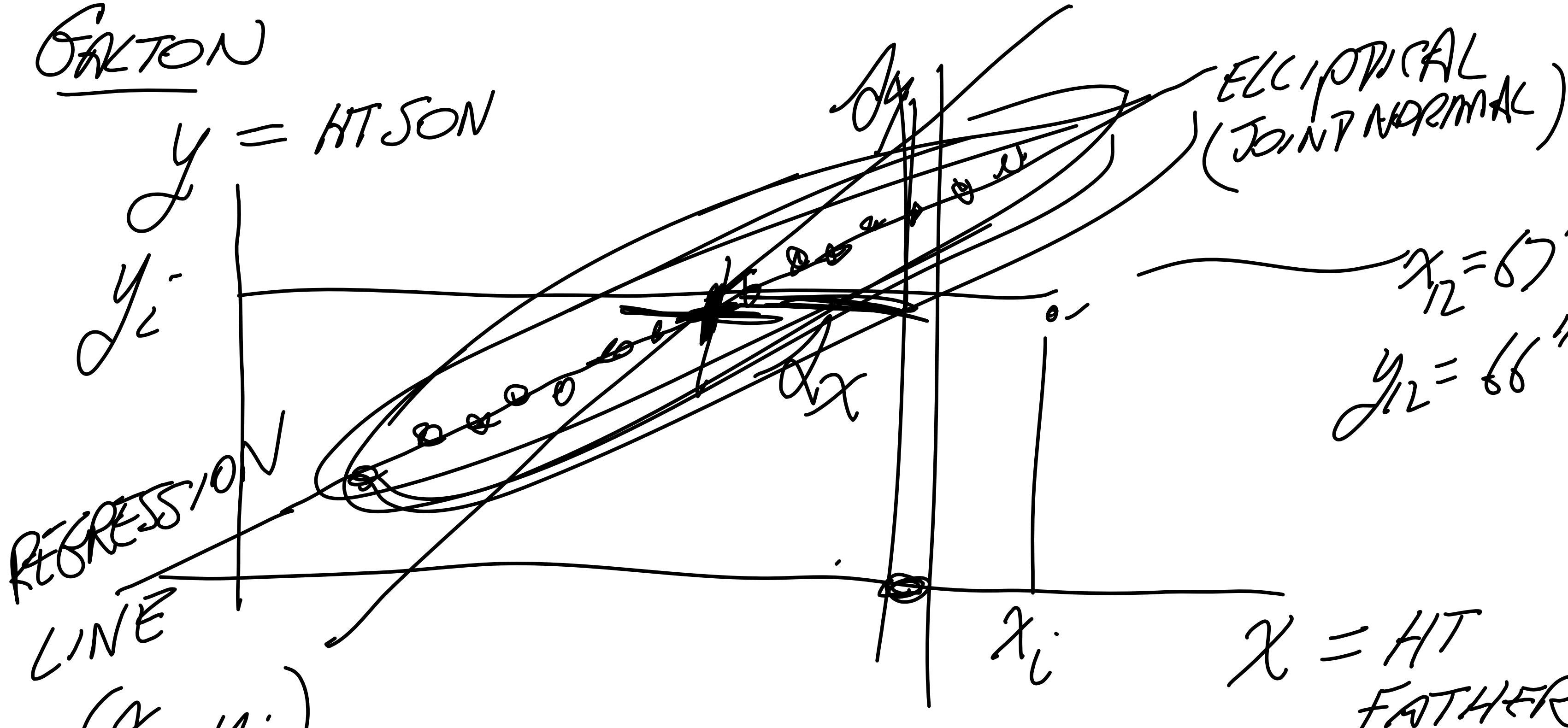
$$\underbrace{\bar{y} + (\mu_x - \bar{x})\, r \frac{s_y}{s_x}}_{\text{IMPROVED (ENHANCED) ESTIMATOR OF } \mu_y} \pm 1.96 \frac{s_y}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \sqrt{1-r^2}$$

SAMPLE CORRELATION $r = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{x}}{s_y}\right) \Big/ (n-1)$

z_{x_i} = (ESTD) STD SCORE OF i^{TH} PERSON'S x-SCORE $= \frac{x_i - \bar{x}}{s_x}$

$$z_{y_i} = \frac{y_i - \bar{x}}{s_y}$$

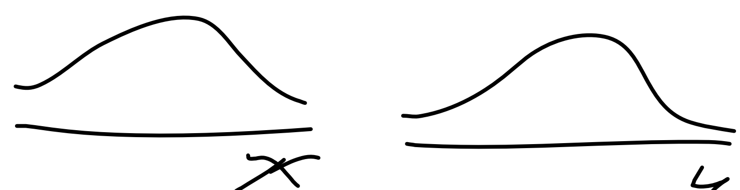
GALTON



(x_i, y_i)

HT OF FATHER i IN SAMPLE

HT OF SON BORN TO FATHER i IN SAMPLE AT MATURITY



REGRESSION TO MEDIOCRITY:

eg AVG HT OF SONS WHOSE FATHERS ARE 1 SD ABOVE AVG HT IS LESS THAN 1 SD OVER AVG.