AMS 7 11/17/09

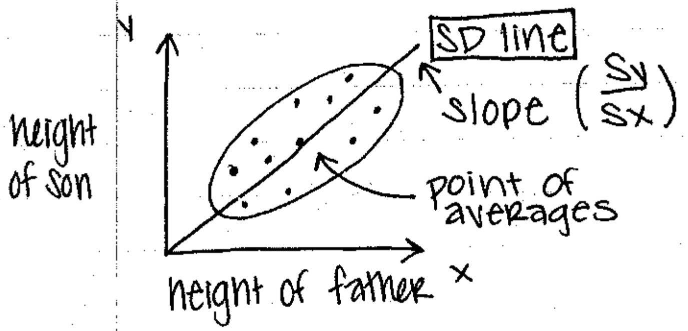
REGRESSION

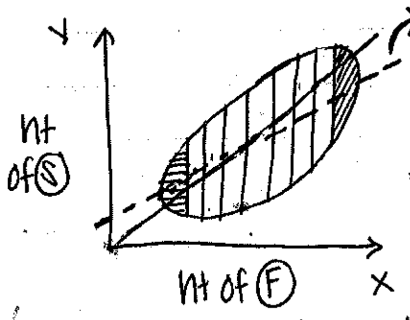
HW 4 due Tue, Nov 24

Lab 6 due Mon, Nov 23

No discussion sections or labs Thanksgiving week (Nov 23-25)

extremes will not follow SD line





mean of vertical strip = best guess for point on line

= REGRESSION LINE

⇓

Conclusion: tall fathers have tall sons, but not as tall as they are. short" "

⇓

data "regresses" towards the mean

• REGRESSION LINE: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

(predicted y) (est. slope)(x)

(est y intercept)

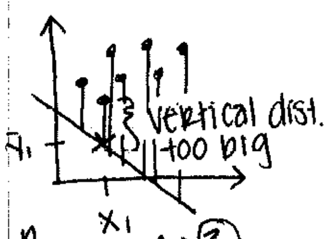
(R-26) $\boxed{\text{MATH FACT}}$: $\boxed{\hat{\beta}_1 = r \cdot \frac{S_y}{S_x}}$

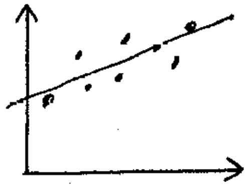
• Sparrows: $(.8704)\left(\frac{0.3499 \text{ cm tail length}}{0.3950 \text{ cm wing length}}\right) = 0.771 \frac{\text{cm tail}}{\text{cm wing}}$

(L-217) JMP : $\hat{\beta}_1$ in PARAMETER ESTIMATES (Lab 6)

(R-26) (17) when $x = \bar{x} \Rightarrow y = \bar{y} \Rightarrow \hat{\beta}_0 + \beta_1 \bar{x} \rightarrow \boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x}}$

(L-246) GAUSS (1800) goal: find EQ of best line of points





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

LEAST SQUARES LINE (LS)

(L-247) $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ = sum of difference between line & pts.

(2) → prevents sum = 0

- REGRESSION LINE = LEAST SQUARES LINE ⇒ INFERENCE IN REGRESSION

L-248
- see sparrow data set

$\boxed{\text{MATH FACTS}}$

$\boxed{1}$ $E_{IID}(\hat{\beta}_1) = \beta_1$

$\boxed{2}$ $\widehat{SE}_{IID}(\hat{\beta}_1) = \dfrac{(S_{Y|X})}{S_X\sqrt{n-1}}$ ← Y given X = RESIDUAL SD

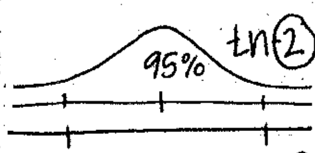
JMP: "root mean square(d) error"

$$(S_{Y|X}) = S_Y\sqrt{1-r^2}\cdot\sqrt{\frac{n-1}{n-2}}$$

R-246
$$\Rightarrow \widehat{SE}_{IID}(\hat{\beta}_1) = \frac{S_Y\sqrt{1-r^2}}{S_X\sqrt{n-2}}$$

$$S_{Y|X} = (0.3499)\sqrt{1-(0.8704)^2}\sqrt{\frac{12-1}{12-2}} = 0.1807 \text{ CMTL}$$

$$\widehat{SE}_{IID}(\hat{\beta}_1) = \frac{0.1807}{(0.3950)\sqrt{12-1}} = .1379\ \frac{\text{CMTL}}{\text{CMWL}}$$

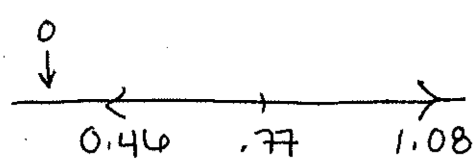


$t_{n-2} \longrightarrow \hat{Y} = \underset{①}{\hat{\beta}_0} + \underset{②}{\hat{\beta}_1} X$

$2.228 \Rightarrow t_{10}^{0.95}$

95% CI = 0.7709 ± (2.228)(0.1379) = (0.46 , 1.08)

L-251



⇒ slope is STAT SIG

HW

Q: Is this practically significant? : Is a correlation sig?
A: yes for the same reasons

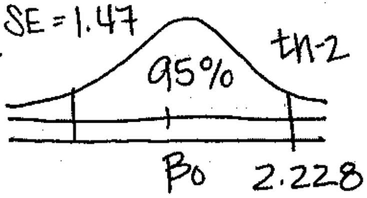
MATH FACTS

1. $E_{IID}(\hat{\beta}_0) = \beta_0$

2. $\hat{SE}_{IID}(\hat{\beta}_0) = \dfrac{S_{Y|X}}{\sqrt{n}} \sqrt{1 + \left(\dfrac{n}{n-1}\right)\left(\dfrac{\bar{x}}{S_*}\right)^2}$

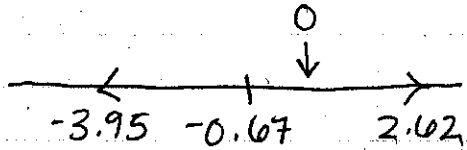
$$\hat{SE}_{IID}(\hat{\beta}_0) = \frac{\overset{(CMTL)}{0.1807}}{\sqrt{12}} \sqrt{1 + \left(\frac{12}{11}\right)\left(\frac{10.683 \text{ WL}}{8.3950 \text{ WL}}\right)^2} = 1.474 \text{ CMTL}$$

3. $\hat{SE} = 1.47$

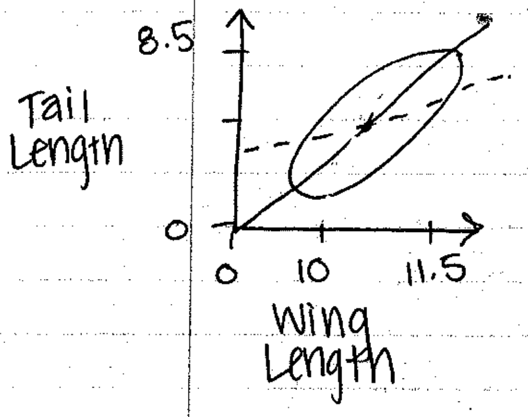


$t_{10}^{0.95} = 2.228$

$-0.669 \pm (2.228)(1.474)$
$= (-3.95, +2.62)$

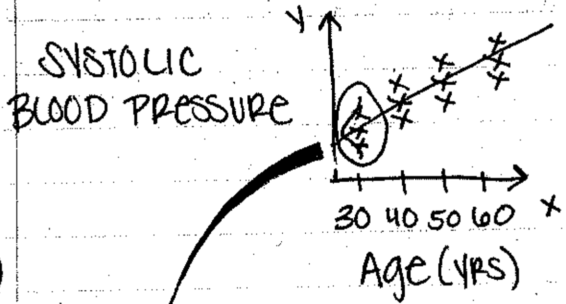


- 0_1 is in the 95% CI as it should b/c as tail length goes down to zero, so should wing length



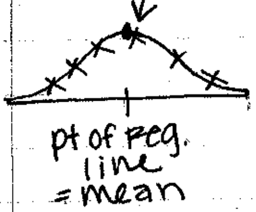
- uncertainty will increase as you get further away from the data
- i.e.: wide uncertainty bands
- EXTRAPALATION of data far away from bulk of data in x is risky

Reg. line = Regression line
pt. = point



L-254

possible bp	bp	age
$\hat{Y}_i$	Y_i	X_i



→ values will equal small normal curve →
center (mean) of normal curve = pt of line

- obs = truth + random error
- $Y_i = (\beta_0 + \beta_1 X_i) + C_i$ $\left\{ \begin{array}{l} (\beta_0, \beta_1; C_i) = \text{unknown} \\ (X_i) = \text{observed} \end{array} \right\}$
- (C_i) = IID normal with mean 0 and SD $\boxed{\sigma_{Y|X}}$ ← residual SD
- obs = truth + error

L-218

- $Y_i = \underbrace{(\hat{\beta}_0 + \hat{\beta}_1 x_i)}_{\text{predicted } (\hat{Y}_i)} + \hat{e}_i$
- $\left[\begin{array}{l} \hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \\ = \text{observed} - \text{predicted} \end{array} \right]$

$\hat{e}_i$ → 

- $\hat{e}_i$ = RESIDUALS ⇒ difference between data pt. & reg. line

L-218

JMP: wing length = x_i tail length = y_i
predicted tail length = $\hat{y}_i$
residuals tail length = $\hat{e}_i$

- $\hat{\sigma}_{Y|X}$ ⇒ the amount by which you expect Y & $\hat{Y}$ to differ

Residual SD: $\hat{\sigma}_{Y/X} = \underbrace{\sqrt{\underbrace{\frac{1}{n-2} \sum_{i=1}^{n} (\hat{e}_i)^2}_{\text{mean sq. error}}}}_{\text{root mean sq. error}}$ ← squared error

L-255

JMP: