June 2, 2017

This time: regression

Next time: ~~?~~ ANOVA

no HW#5

HW#4 is due Friday June 9, 2017 @ 11:59 on canvas

There will be either 1 or 2 makeup lectures in finals week, on Monday (+ Tuesday) more details soon

Take home final given out Friday June 9, 2017 during class and will be due a week later

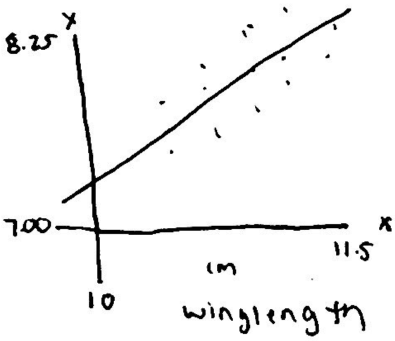
Q_1: when is a regression slope $\hat{\beta}_1$ large in practical terms (practsig)?

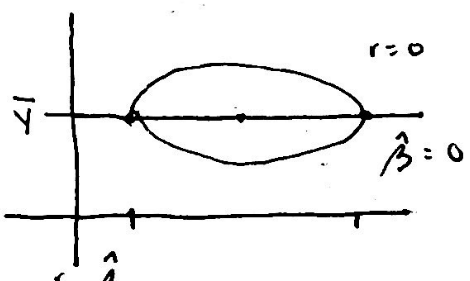
A_1: Same method as in answering the question

Q_2: when is ~~co~~ a correlation coefficient in practical terms

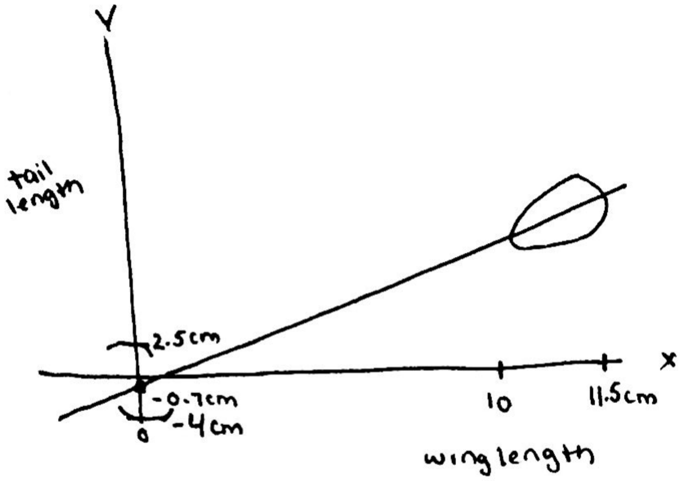
$A_2 = A_1$

smallest birds have x = wing length = 10 cm, $\hat{y}$ = predicted tail length = 7.0 cm; longest have $x = 11.5$, $\hat{y} = 8.25$ cm. Since 8.25cm differs from 7.0 cm by an amount thats practsig, the correlation r + slope $\hat{\beta}_1$ associated w/ predicting y from x are also ~~practsig~~ practsig.



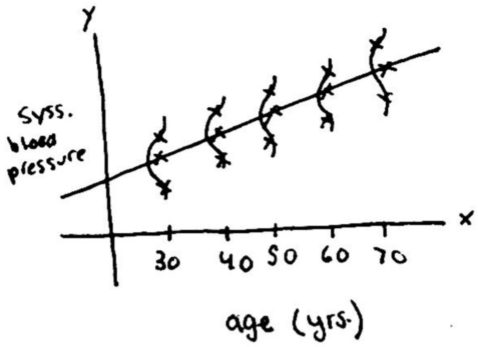


r, $\hat{\beta}_1$ not practsig



often, y-int. involves extropolation away from data; it therefore may be meaningless

Another way



$\cancel{y_i = \beta_0 + \beta_1 x_i) + e_i}$

$y_i = \underbrace{\beta_0 + \beta_1 x_i} + e_i$

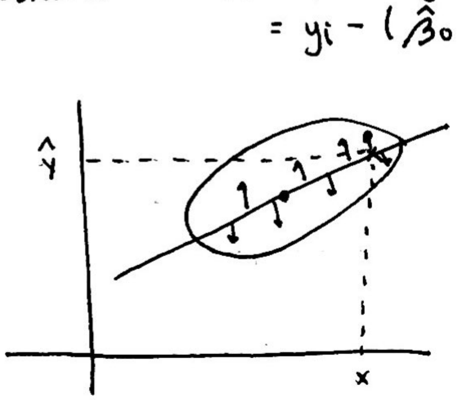
obs = "truth" + "error"

$y_i = (\hat{\beta}_0 + \hat{\beta}_1 x_i) + \hat{e}_i$

obs = $\underset{y}{\text{predicted}}$ + residual

residual $\hat{e}_i = y_i - (\hat{y}_i)$

$= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$



$\hat{y} = \beta_0 + \hat{\beta}_1 x$

If available,

$\hat{Y}_x = \hat{\beta}_0 + \hat{\beta}_1 x$ +

$\hat{SE}(\hat{y}_x) = S_{y|x}$

residuals

= RMSE :

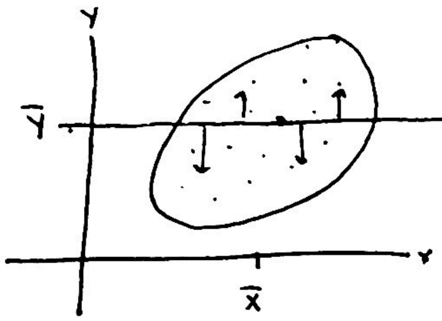
% of variance in y is "explained by" the regression of y on x

change to "associated with" (better)

r^2 = <u>coefficient of determination</u>

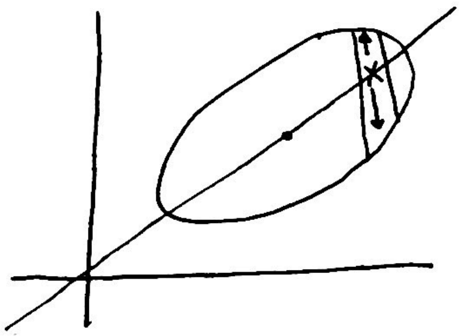
big r^2 = useful
↳ big enough to matter

method 2 ↓



case Ⓐ do not have x or choose to ignore x

best $\hat{y} = \bar{y}$ + $SE(\hat{y}) = S_y$



case Ⓑ use x to predict y

now best $\bar{y} = \hat{\beta}_0 + \beta_1 x$
and $\hat{SE}(\hat{y}) = S_{y|x}$
$= \sqrt{1-r^2}$

regression is successful if $S_y\sqrt{1-r^2}$ is smaller than S_y by an amount that's ~~meanful~~ meaningful

One way to quantify this is

$$\frac{\underset{\text{ignore } x}{S_y} - \underset{\text{use } x}{S_y\sqrt{1-r^2}}}{\underset{\text{ignore } x}{S_y}} = 1-\sqrt{1-r^2}$$

beter measure of success of regression than r^2