

May 26, 2017

This time: correlation + regression
Next time: ↓ ↓

read: LN pp. L-214 → L-268

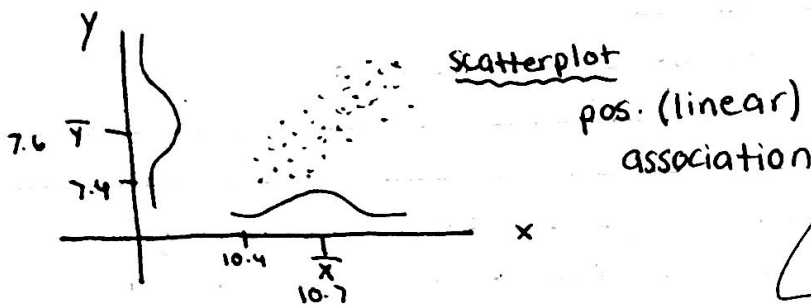
HW #3 due tonight @ 11:59pm on Canvas

No class or discussion on Monday, those who have discussion that day must go to another one (Tue-Fri)

Right Now you can do #1 + #2 on HW #4 → expected due date is Friday June 2nd

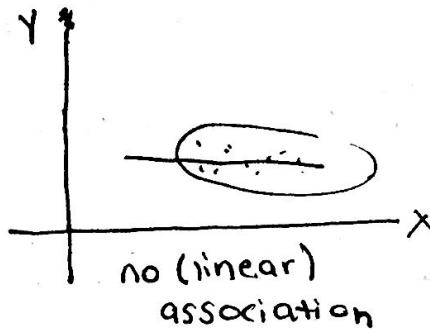
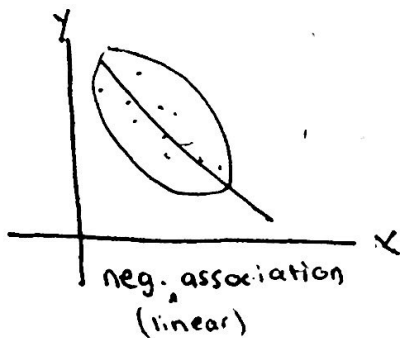
Case Study

- wing lengths of sparrows

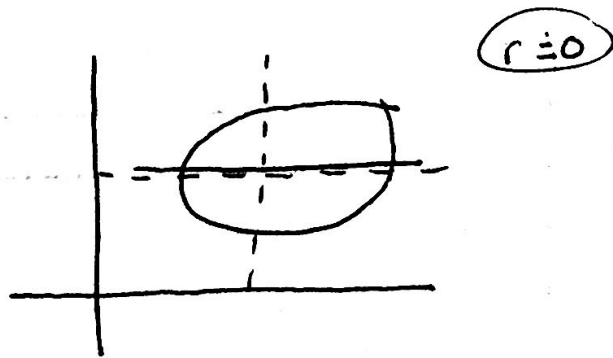
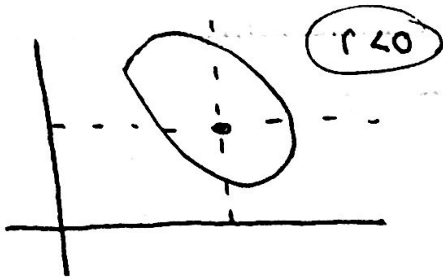
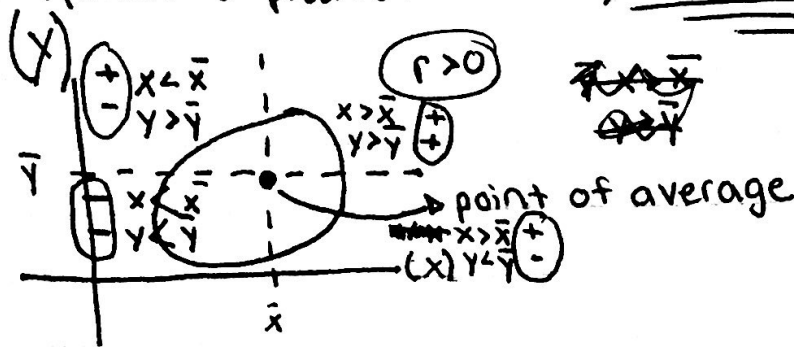


Karl Pearson +
Francis Galton 1885

basic shape of a scatterplot is an ellipse ~~ellipse~~
elliptical (bivariate normal)



$r =$ (pearson's product moment) correlation coefficient



	Y	X
	y_1	x_1
	\vdots	\vdots
	y_2	x_2
	\vdots	\vdots
	y_n	x_n
mean	\bar{y}	\bar{x}
SD	S_y	S_x

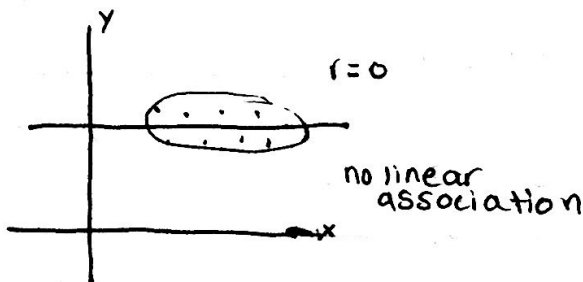
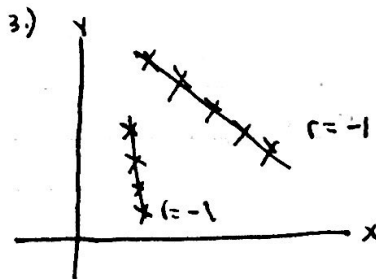
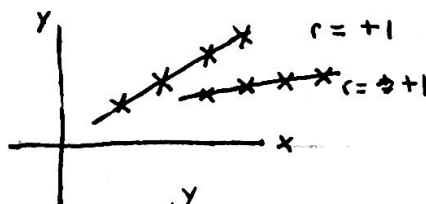
$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \bar{y}}{S_y} \right) \cdot \left(\frac{x_i - \bar{x}}{S_x} \right)$$

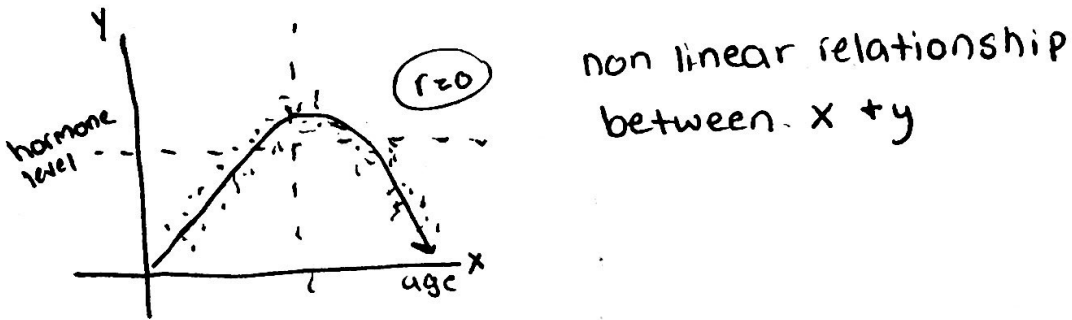
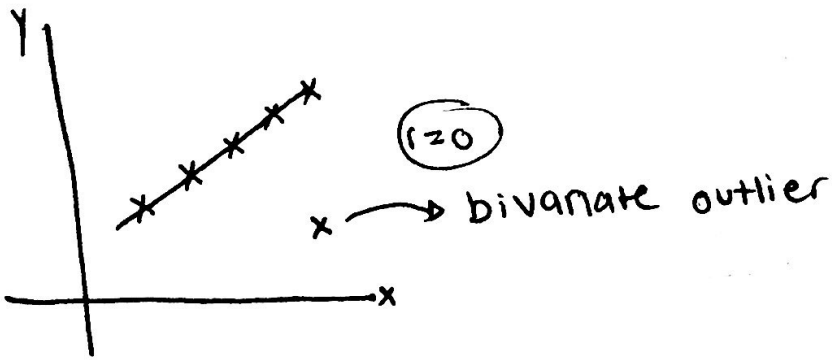
$$S_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$S_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

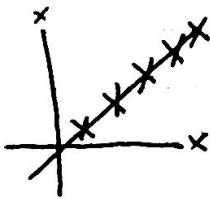
facts about r

- 1.) r is a pure number, without units
- 2.) $-1 \leq r \leq 1$



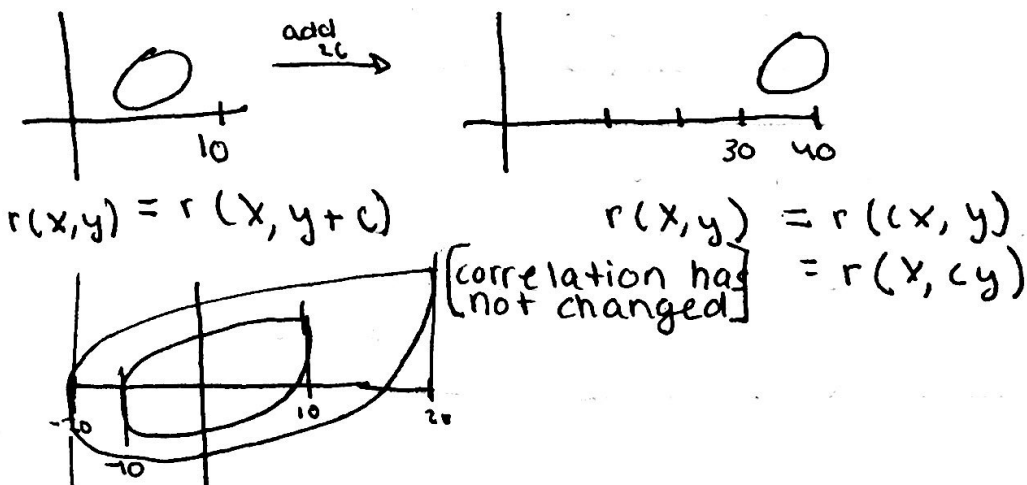


4.) $r(x, x) = +1$ for any variable x
 ↳ correlation ~~bet~~ between x and x



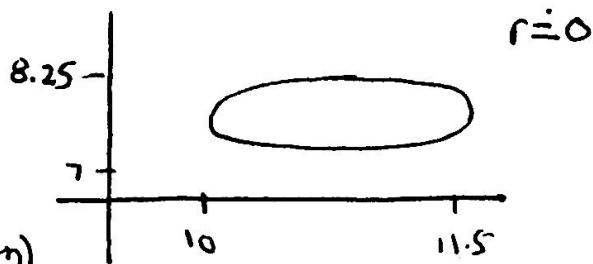
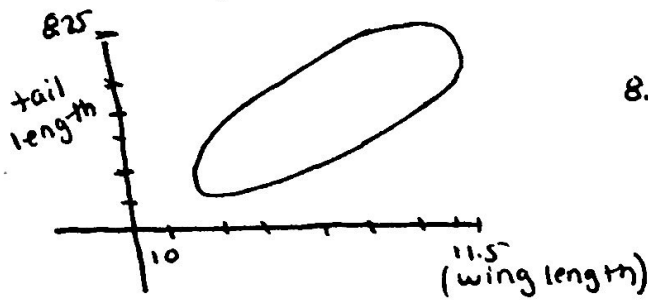
5.) $r(x, y) = r(y, x)$
 for all x var. $x + y$

6.) add C to all x values $\rightarrow r(x, y)$



Q: is $r(\text{wing length, tail length}) = +0.87$ large in practical terms?

* Does ~~0.87~~ 0.87 differ from 0 by an amount that matters (real-world)?
 null boring value $r=0$



as we go from one sparrow w/ $(x=10, y=7)$ to $(x=11.5, y=8.25)$; 8.25 is much larger practically than 7; but if $r=0$, $(x=10, y=7.6)$ $(x=11.5, y=7.6)$ no diff.

model drawn on p L-249

