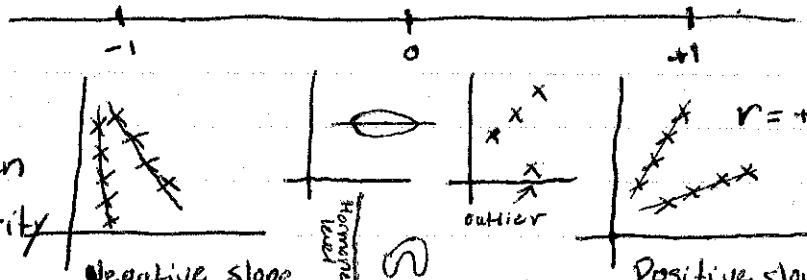
AMS 2
Notes
1
11/12
Julia Larson

11/12 Correlation facts about $r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x^*}\right)\left(\frac{y_i - \bar{y}}{s_y^*}\right)$

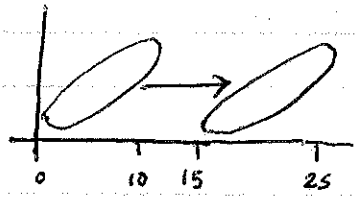
- r is a pure number without units
- r is always $-1 \leq r \leq +1$



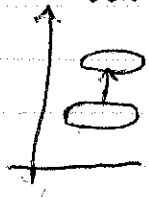
- r can be fooled by (a) outliers (especially when n is small) or b nonlinearity or both. Ask to see a scatter plot if r is close to 0.

x's lie on a perfect line

- if I add a constant to all the x values
  exp.: add 15 to x

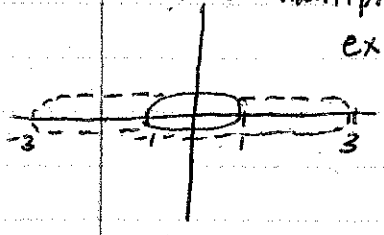


add constant to Y



R-98 - Train to see correlations in scatterplots

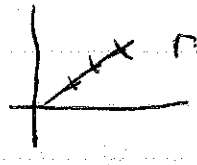
- If I multiply all x (or y) values by a positive constant:
  exp. multiply x by 3



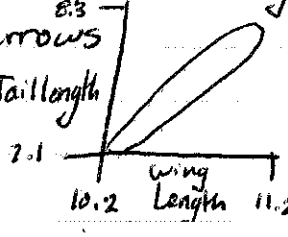
$$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x^*}\right)\left(\frac{y_i - \bar{y}}{s_y^*}\right) \quad \frac{(3x_i - 3\bar{x})}{3s_x^*}$$

** is unchanged

- The correlation between a variable and itself. If I interchange the role of x and y. r in unchanged

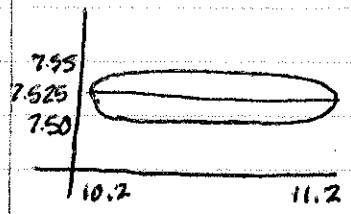


- Practical significance: The correlation between x and y is large in practical terms when knowledge of x definentely helps you to predict Y example: Sparrows



- when x is around 10.2 Y is around 7.1, when x is around 11.2, Y is around 8.3; and 7.1 and 8.3 definentely differ by a amount that's large in practical terms, so is pract sig.

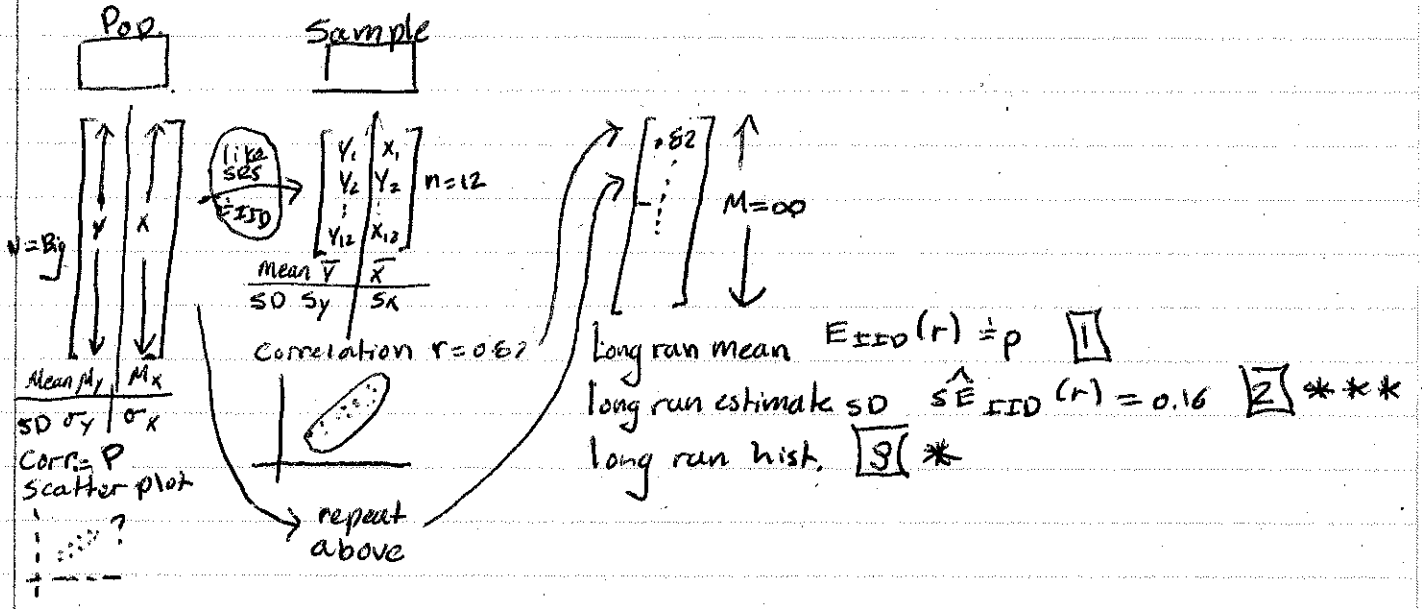
Notes
AMS 7
2
11/12
Julia Larson



Not pract sig. not a lot of difference between 7.50 and 7.55 (N)as you go along x-axis

Inferential Summary

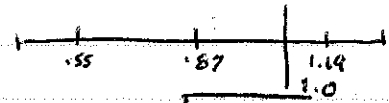
Unknown pop. quantity of Interest	ρ = pop. Corr between wing and tail length
estimate of	$r \doteq +.67$
give or take	$\hat{SE}(r) \doteq .16$
95% interval	Approx (.55, 1.19)



1 $E_{IID}(r) \doteq \rho$

2 $\hat{SE}_{IID}(r)$ we want n↑ SE↓ $= \sqrt{\frac{1-(+.67)^2}{12-2}} = 0.16$

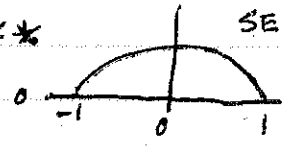
3 Long run hist. of r (try) normal curve fails numbers come out more extreme than +1 or −1, so truncate.



$r \pm 1.96\hat{SE}(r)$

$.87 \pm 2(0.16)$

***



$SE(r)$ = Math Fact $SE_{IID}(r) \doteq \sqrt{\frac{1-\rho^2}{n-2}}$

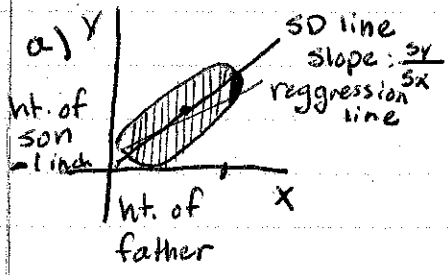
$\hat{SE}_{IID}(r) = \sqrt{\frac{1-r^2}{n-2}}$

To get real 3 x-credit L 231 - 244

0 is outside int, so corr. is statsig.

AMS 7
Notes
3
11/12
Julia Larson

Regression: What's the equation of line capturing trend of ellipse; turns out to be 3 lines of interest in any ellipse: best line for
a) predicting y from x
b) predicting x from y
c) capturing overall trend. SD line



- Goal to predict Y from X
- Galton (1860's - 1905's)
- This line should go through $(\bar{x}, \bar{y})$