this: 2 indep time: samples, 0/1 outcomes

next time: correlation

read: LN 19.2L-214 → L-248 | AMS7 22 May 18

today: LN 19.L-250 →

$$\widehat{SE}_{IID}(\hat{p}_1) = \frac{\sigma_1}{\sqrt{n_1}} = \frac{\sqrt{p_1(1-p_1)}}{\sqrt{n_1}}$$

$$\widehat{SE}_{IID}(\hat{p}_1) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}} = \sqrt{\frac{p_1(1-p_1)}{n_1}}$$

$= 1.1\%$

$$\widehat{SE}_{IID}(\hat{p}_2) = \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 1.5\%$$

$$\widehat{SE}_{\text{IID, indep.}}(\hat{p}_1 - \hat{p}_2) = ?$$

≥ 1.5, ≤ 3.0

$\widehat{SE}(\hat{p}_1 - \hat{p}_2)$

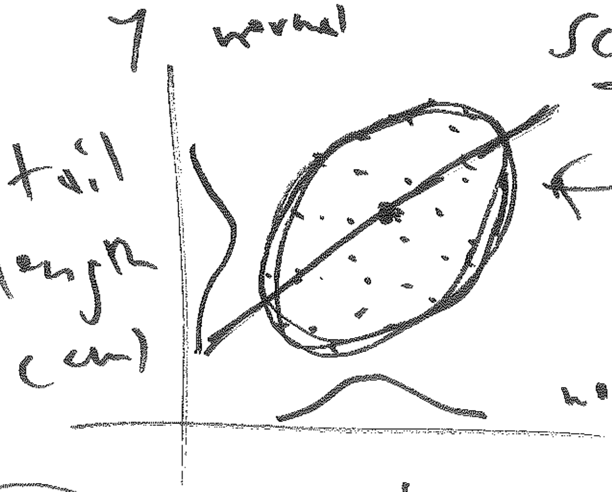
$\widehat{SE}(\hat{p}_1) = 1.5$

$\widehat{SE}(\hat{p}_1)$

1.1

$$\widehat{SE}_{\text{IID, indep}}(\hat{p}_1 - \hat{p}_2) = \sqrt{\left(\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}\right)^2 + \left(\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)^2}$$

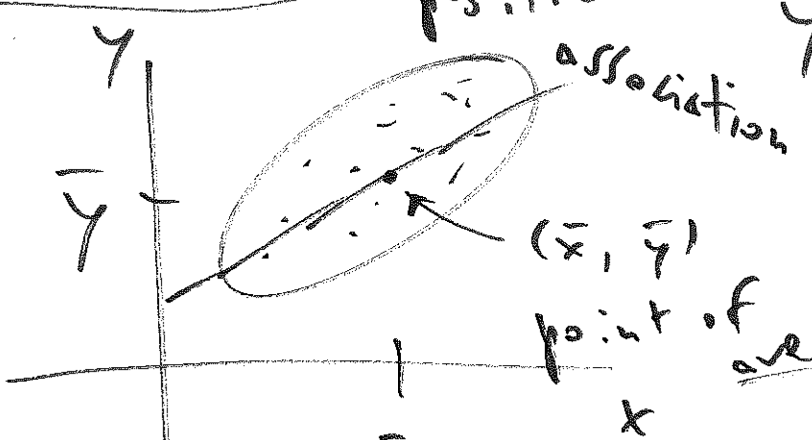
$$\widehat{SE}_{\text{IID, indep}}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

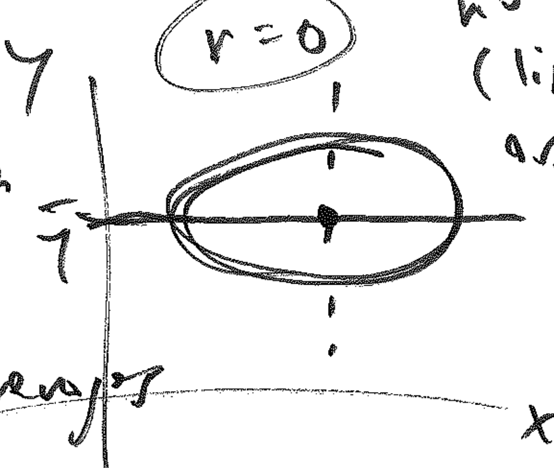


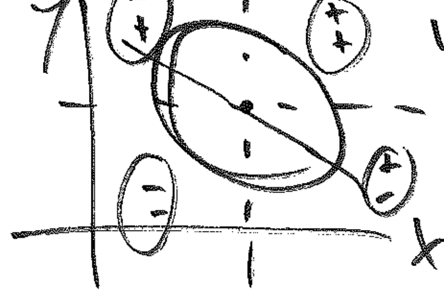
scatterplot (scatter diagram)

ellipse (summary of a bivariate normal distribution)

1890s
Karl Pearson
Francis Galton

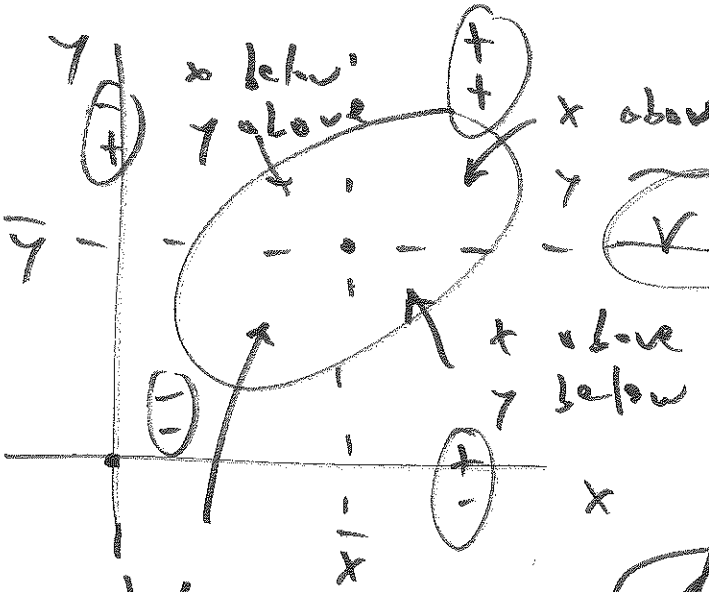






negative association

$r < 0$



$$\frac{x - \bar{x}}{s_x} > 0 \qquad \frac{y - \bar{y}}{s_y} > 0$$

$r > 0$

$$\frac{x - \bar{x}}{s_x} < 0 \qquad \frac{y - \bar{y}}{s_y} < 0$$

$r > 0$

sample correlation coefficient between x & y

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

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

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

y	x
y_1	x_1
y_2	x_2
$\vdots$	$\vdots$
y_n	x_n

n

mean $\bar{y}$ $\bar{x}$

SD s_y s_x

facts about r

① $-1 \le r \le +1$