

Panel 1

Last Time:

See "Contingency worksheet" for review on contingency tables and Chi-Square test of independence.

Apply to Nominal, Ordinal values. For numeric vars., learn in another way: Linear Regression

Linear Regression

- ① Visualize data
- ② "Best-fit" line
- ③ Compute correlation coefficient
- ④ Make predictions.

1

Panel 2

Formula:

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

2

Panel 3

Highest year of school completed, father x	Highest year of school completed y	x^2	y^2	xy	
12	12	144	144	144	12-12
15	16	225	256	240	15-16
5	7	25	49	35	
16	19	256	361	304	
$\Sigma x = 48$	$\Sigma y = 54$	$\Sigma x^2 = 650$	$\Sigma y^2 = 810$	$\Sigma xy = 723$	

compute the correlation coefficient r and the equation of the least-square regression line $y = mx + b$. Responding formulas:

$$s_x = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{650 - \frac{48^2}{4}}{4-1}} = \sqrt{\frac{650 - 576}{3}} = \sqrt{24.67} \approx 4.97$$

$$s_y = \sqrt{\frac{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}{n-1}} = \sqrt{\frac{810 - \frac{54^2}{4}}{4-1}} = \sqrt{\frac{810 - 729}{3}} = \sqrt{28.33} \approx 5.32$$

$$r = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\sqrt{(\Sigma x^2 - \frac{(\Sigma x)^2}{n})(\Sigma y^2 - \frac{(\Sigma y)^2}{n})}} = \frac{723 - \frac{48 \cdot 54}{4}}{\sqrt{(650 - \frac{48^2}{4})(810 - \frac{54^2}{4})}} = \frac{723 - 648}{\sqrt{74.81}} = \frac{75}{\sqrt{74.81}} \approx 0.97$$

$\Sigma x^2 = 650$, $\Sigma xy = 723$, $\Sigma y^2 = 810$

Panel 4

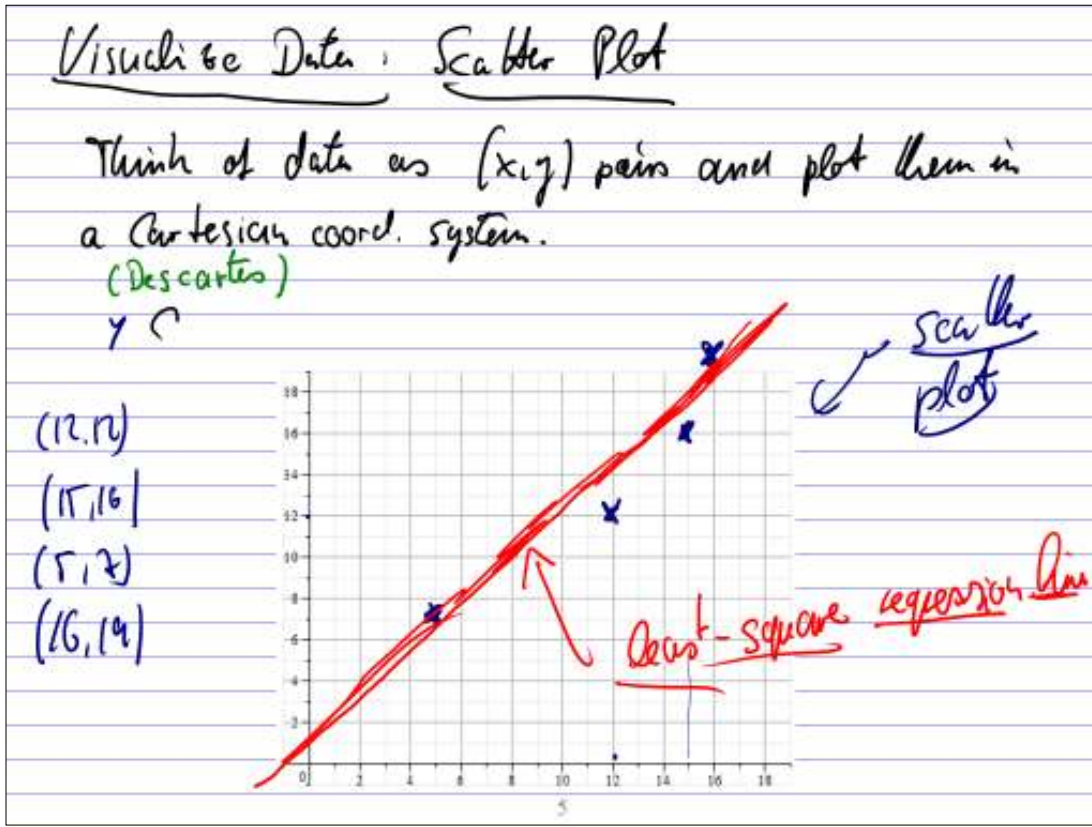
Correlation Coefficient

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

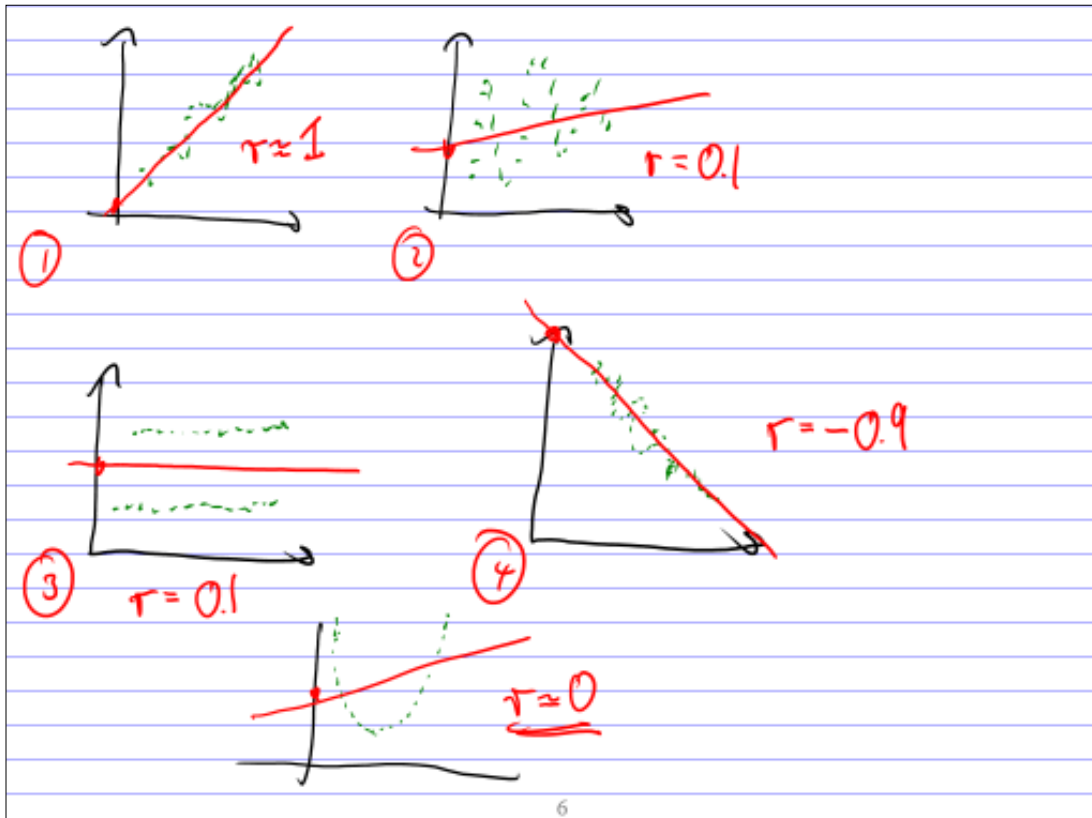
- is always between -1 and 1
- if close to +1 or -1 then close relation between vars
if close to 0, no relation
 - if close to +1: pos. relation (more $x \Rightarrow$ more y)
 - if close to -1: neg. rel. (more $x \Rightarrow$ fewer y)

In example: $r = 0.97 \Rightarrow$ strong pos. relation!

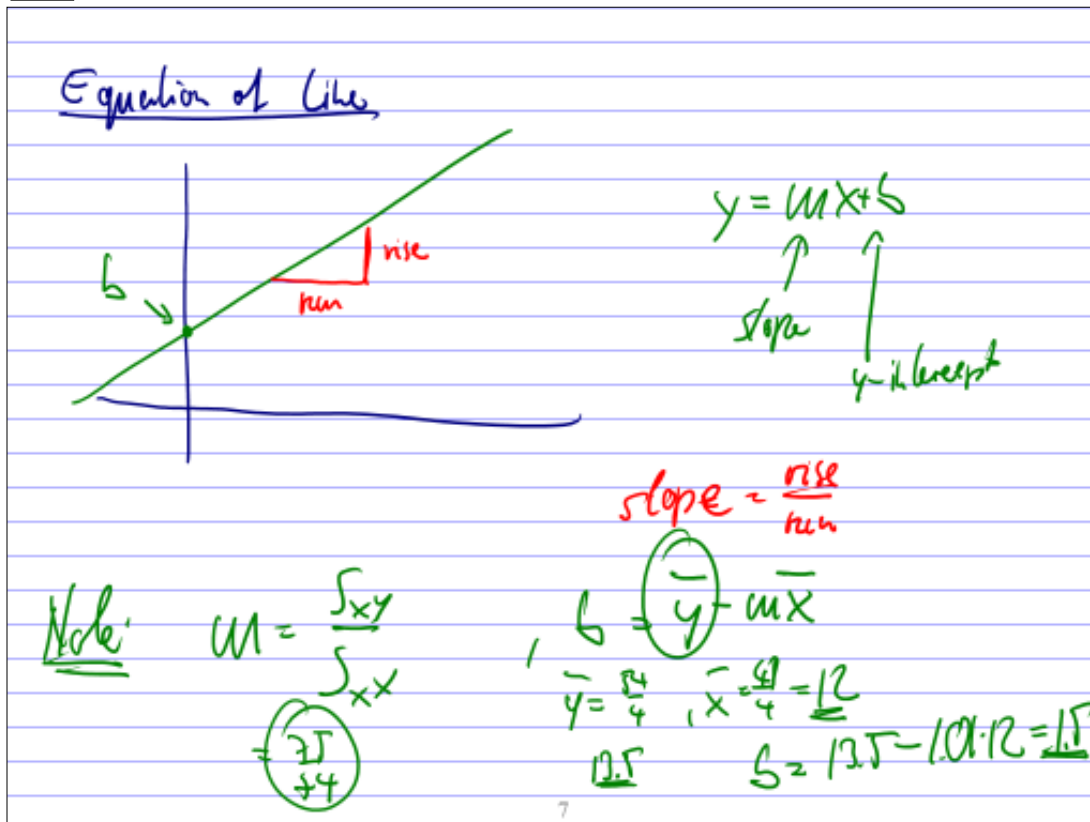
Panel 5



Panel 6



Panel 7



Panel 8

