

10. Linear Regression

p1

We'll be looking at bivariate data (2 variables):

X	Y

Association = Relationship between X and Y

Scatterplots:

a)

resale
value
(\$)



age of car

Negative Association

b)

grade



study time

Positive Association

c)

grade



student's height

Zero Association

d)

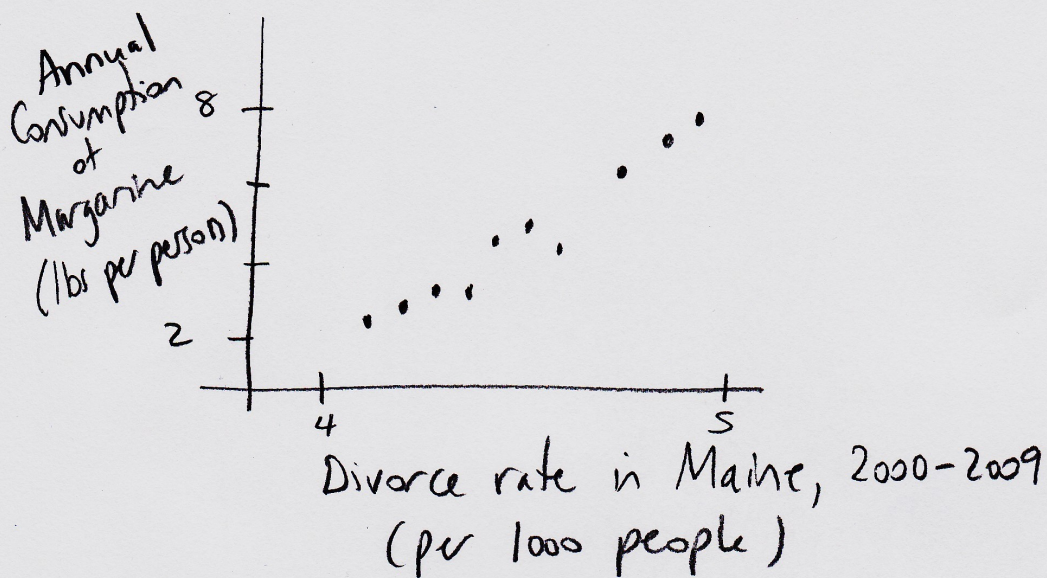
quantity
sold



price of tablet

Nonlinear Association

Correlation does not imply
causation !!!

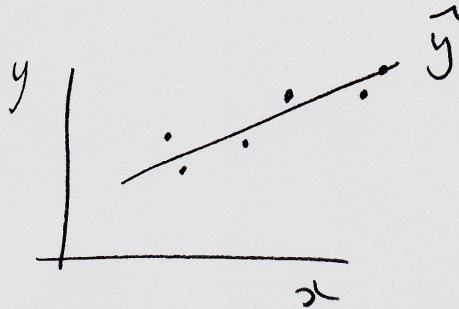


$$r \approx 0.9926$$

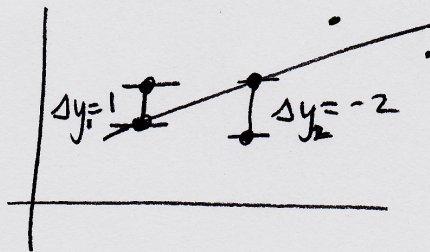
More divorces do not cause
more margarine-eating

Least Squares Regression Line (Best Fit Line)

written $\hat{y}$



Residual of a point
 $\Delta y = y - \hat{y}$



Least Squares Regression Line
minimizes $(\Delta y_1)^2 + (\Delta y_2)^2 + \dots$
summed over all points

p5

Coefficient of determination r^2 :
indicates what % of the variation in y
is accounted for by the best-fit line

Ex: A bivariate data set has $r = -0.8$
 $r^2 = 0.64$

The best-fit line accounts for 64%
of the variation in y

Ex:

$x = \text{age of Toyota Grolla}$ (years)	$y = \text{resale value}$ (\$1,000)
2	5.4
3	5.1
5	4.9
7	4.8
10	4.2

The best-fit line is $y = -0.13x + 5.61$
The coefficient of determination is 0.9522

a) Is the association positive or negative?

NEGATIVE As $x \uparrow$, $y \downarrow$

b) Find the Correlation Coefficient

p6

$$r^2 = 0.9522$$

$$r = \pm \sqrt{0.9522}$$

negative association

$$r \approx -0.98$$

c) What % of the variation in y is accounted for by the best-fit line?

$$r^2 = 0.9522$$

$$\boxed{95.22\%}$$

d) What resale value is predicted for a 4-year-old car?

$$x = 4 \rightarrow y = -0.13x + 5.61$$

$$y = 5.09$$

$$\boxed{\$5090}$$

e) Why should we not predict the resale value for a 1-year-old car?

We should not predict for values outside data set ($2 \leq x \leq 10$)

f) What age corresponds to a resale value of \$4500?

p7

$$y = 4.5 \rightarrow 4.5 = -0.13x + 5.61$$

$$-1.11 = -0.13x$$

$$x \approx 8.5 \text{ years}$$

g) Interpret the slope of $\hat{y}$

$$\text{slope} = -0.13$$

As x increases by 1, y decreases by 0.13 on average

As the car ages 1 year, resale value decreases by \$130, on average.

To calculate $\hat{y}$ and r^2

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

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

$$b = \frac{S_{xy}}{S_{xx}}$$

$$a = \bar{y} - b\bar{x}$$

$$\hat{y} = a + bx$$

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

$$r^2 = \frac{(S_{xy})^2}{S_{xx} S_{yy}}$$

Ex: Find $\hat{y}$ and r^2

p8

x	y
2	5
8	4
9	2

	x	y	x^2	xy	y^2
	2	5	4	10	25
	8	4	64	32	16
	9	2	81	18	4
sum	19	11	149	60	45
mean	$\frac{19}{3}$	$\frac{11}{3}$			

$$\begin{aligned} S_{xy} &= 60 - \frac{19 \cdot 11}{3} \\ &= \frac{180}{3} - \frac{209}{3} \\ &= \frac{-29}{3} \end{aligned}$$

$$\begin{aligned} S_{xx} &= 149 - \frac{19^2}{3} \\ &= \frac{447}{3} - \frac{361}{3} \\ &= \frac{86}{3} \end{aligned}$$

$$\begin{aligned} b &= \frac{S_{xy}}{S_{xx}} \\ &= \left(\frac{-29}{3} \right) \\ &= \frac{\left(\frac{86}{3} \right)}{\left(\frac{86}{3} \right)} \\ &= \frac{-29}{3} \cdot \frac{3}{86} \\ &= -\frac{29}{86} \end{aligned}$$

$$\begin{aligned} a &= \bar{y} - b\bar{x} \\ &= \frac{11}{3} + \frac{29}{86} \left(\frac{19}{3} \right) \\ &= \frac{11(86)}{3(86)} + \frac{29(19)}{3(86)} \\ &= \frac{1497}{258} \end{aligned}$$

$$\hat{y} = a + bx$$

$$\boxed{\hat{y} = \frac{1497}{258} - \frac{29}{86} x}$$

p10

$$\begin{aligned} S_{yy} &= 45 - \frac{(11)^2}{3} \\ &= \frac{135}{3} - \frac{121}{3} \\ &= \frac{14}{3} \end{aligned}$$

$$\begin{aligned} r^2 &= \frac{(S_{xy})^2}{S_{xx} S_{yy}} \\ &= \frac{\left(-\frac{29}{3}\right)^2}{\left(\frac{86}{3}\right)\left(\frac{14}{3}\right)} \quad \times \quad \frac{9}{9} \\ &= \frac{29^2}{86(14)} \\ &= \frac{841}{1204} \end{aligned}$$

Compute $\hat{y}$ and r on calculator

p11

MODE 1 1

2 STO 5 M+

8 STO 4 M+

9 STO 2 M+

RCL r

$$r \approx -0.84$$

RCL a

$$a \approx 5.80$$

RCL b

$$b \approx -0.34$$

$$\hat{y} = a + bx$$

$$\hat{y} \approx 5.80 - 0.34x$$