

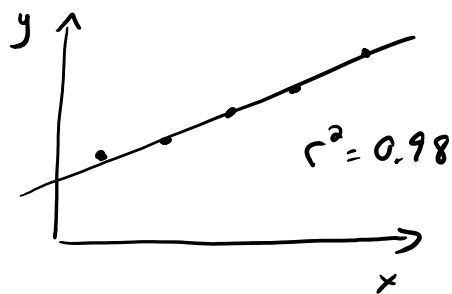
r^2 is called the coefficient of determination

-why would we bother?

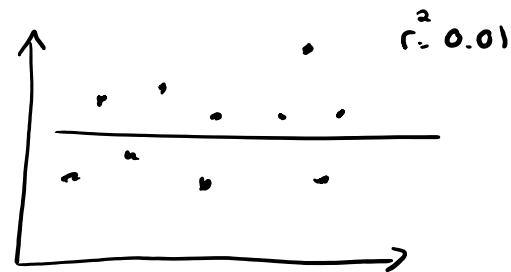
r^2 indicates what % of the variation in y is accounted for by the best-fit line

example: If a bivariate data set has $r = -0.8$, then $r^2 = 0.64$ and so the best-fit accounts for 64% of the variation in y

what does this mean?



2% is due to random behaviour and 98% to linearity



99% is due to randomness

①

$$\hat{y} = -0.13x + 5.61$$

\hat{y} is the best fit line for

x : age of a Corolla in years

y : resale value in \$1000s

years	in \$1000s
2	5.4
3	5.1
5	4.9
7	4.8
10	4.2

coeff of determination = 0.9522

r^2

a) is the linear association + or - ?

- (negative slope, also as x increases, y decreases)

b) find the correlation coefficient

$$r^2 = 0.9522$$

$$r = \pm \sqrt{r^2} = -0.9758$$

↑
from negative association

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

95.22%

d) what resale value is predicted for a 4-year-old Corolla?

