

DS2000

11/26 - Tues :)

Admin

- Please do trace!

- 12/3 last lecture: bit.ly/49ann1T

Agenda

1. Correlation
2. Linear Regression
3. Python

relationship b/w attributes
of TVshow objects

⊗ guesses...

trope vs. rating

tone vs. length

#eps vs. rating

prediction model

9. List Comprehensions

↳ shortcut for loop + appending to a list

(never actually needed!)

but, we need to be conversant in them

goze: iterate over
TVshow objs,
make a list
of lengths

long version

lst = []

for show in shows:

lst.append(show.length)

comprehension version

lst = [show.length for show in shows]

- no append

- start []

- each element (show.length)

- last: loop!

1. Correlation

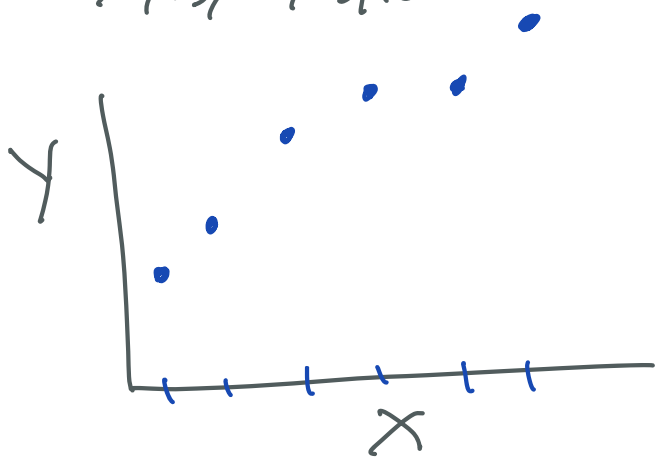
↳ describe the relationship between two variables (lists)

↳ if a strong relationship between two vars, then we can make a Linear regression

↳ make predictions

(ex) $X = 1, 2, 3, 4, 5, 6$

$Y = 5, 9, 13, 15, 15, 18$



plot X vs. Y

not quite a perfect line, but:

when X goes up, Y goes up too

↳ Strong Correlation

Prediction model

given an x I've never seen,
what would y be?

X - independent

Y - dependent

need correlation before we
make the prediction

Statistics. Correlation (X, Y)



closer to -1 : neg correlation

$+1$: pos correlation

0 : no correlation

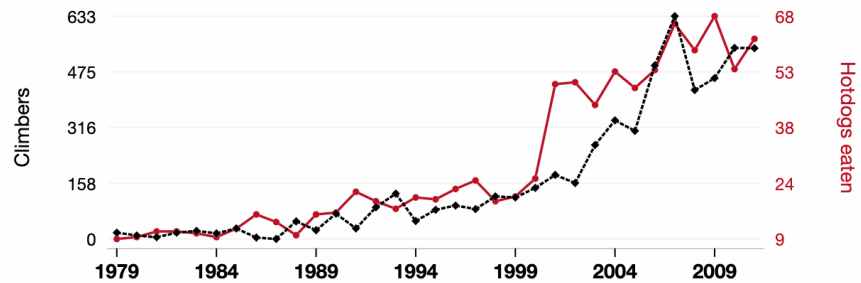
Correlation vs. Causation

- trope vs. rating
- episodes vs. rating
- time vs. length

Total Number of Successful Mount Everest Climbs

correlates with

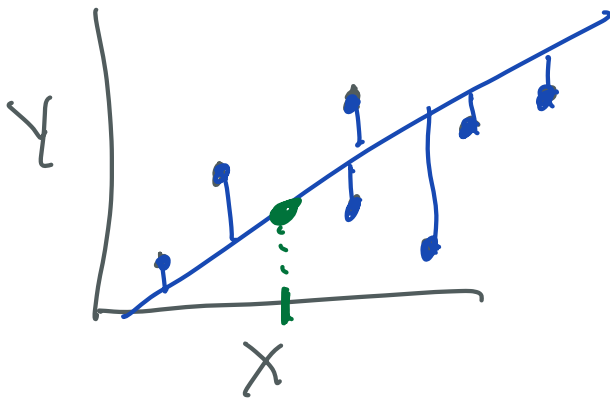
Hotdogs consumed by Nathan's Hot Dog Eating Competition Champion



2. Linear regression

↳ create a line of best fit
prediction model:

choose x on line, not on dataset
predict what y would be



- minimize the distance
between each point
and the line

x never seen
 y on the line

In python: generate an equation for line

$$y = mx + b$$

visualize the line of best fit

`sns.regplot()`

Python

install: statistics, seaborn

`import statistics`

`import seaborn as sn`