This time: descriptive methods

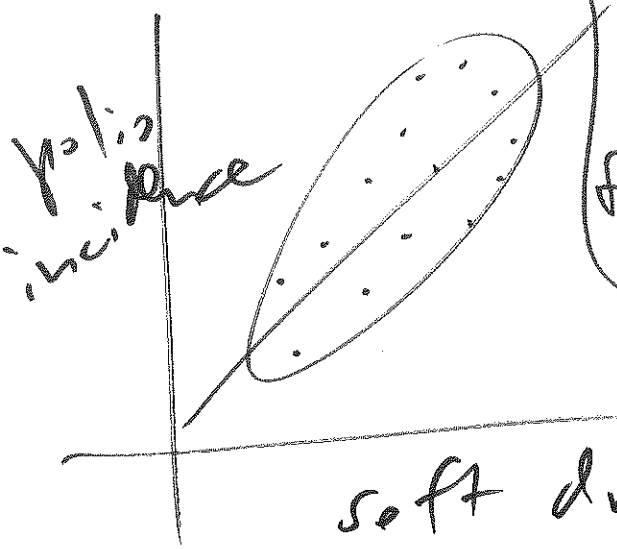
Next time:

season	soft drink	polio
F	M	M
W	L	L
Sp	M	M
Su	H	H

AMS 5 Apr 14

①

H high
L low
M medium



strong positive association

but doesn't prove causality

pop.

← variables →

N = population size

subjects (deer)

population: all deer living at UCSC on April 2017

1 row for each deer

sample: the observed deer

idea: (2)

$\hat{p} = \hat{\theta} = \bar{y}$ is a good estimate of $p = \theta$

1 = Y
0 = N

$N = 800$

disease?
no/N/0
no/N/0
no/N/0
yes/Y/1
no/N/0
⋮
no/N/0

at random (IID)

disease?
1's
2
0's

$n = 85$

$\frac{3}{85} = 3.5\%$ estimate

mean $\bar{y} = \hat{p} = \hat{\theta}$

sample size

$\bar{y}$-bar $\hat{p}$-hat $\hat{\theta}$ theta-hat

mean $p = \theta = ?$

proportion in pop. with disease / unknown

goal of sampling:

representativeness:

mean of 1's and 0's

keeps track automatically of % of 1's

want samples unsample similar in all relevant ways

population

unsample | sample

at random with replacement: IID ← independent identically distributed sampling

at random without replacement: SRS simple random sampling

SRS is more informative than IID, but IID has easier math

---

if n is a lot smaller than N,

SRS $\doteq$ IID

$\doteq$ is approximately equal to