STAT 131

7 Apr 20

~~lecture~~ 1

read: DS ch. 1

this set time: theory, foundations

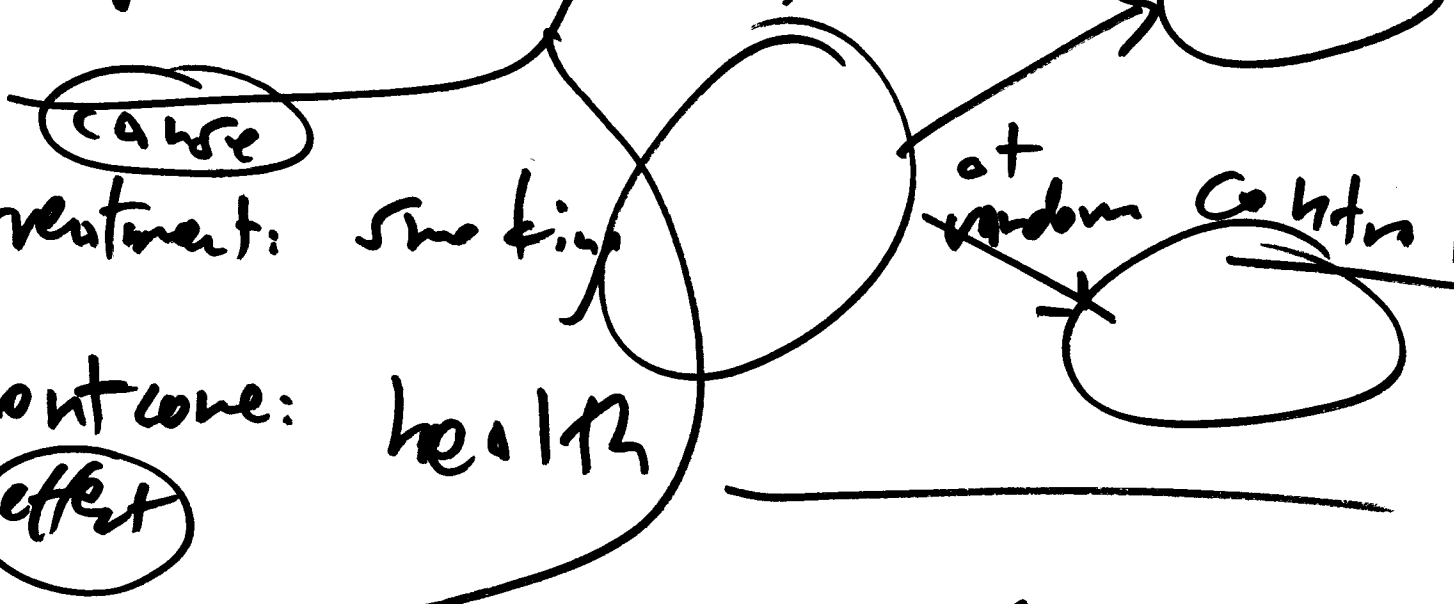
next time: combinatorics

strongest evidence for correct

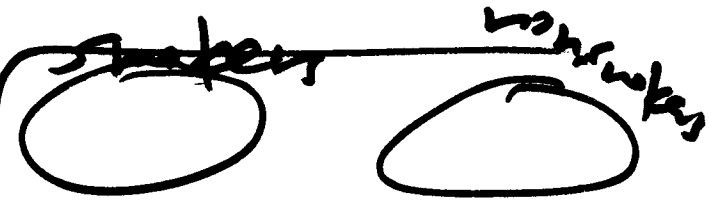
cause → effect

conclusions:

randomized controlled experiment (trial) (RCT)



can't always run RCT on people (ethics)



observational study

health

health

bias: systematic tendency to over- or under-estimate the truth

---

obs. studies can have (lots of) bias because of potential confounding factors (PCFs):

treatment (X): smoking "T" smokers "C" non smokers

outcome (Y): health (heart disease)

PCF (Z): exercise

~~constitutional hypothesis~~

$P(\text{both} \mid CH) = P(HH \mid \text{IID fair coin tossing})$

my cancer smokers

heads

$= P(\text{H 1st and H 2nd} \mid \sim)$

$= P(\text{H 1st} \mid \sim) \; P(\text{H 2nd toss} \mid \text{H 1st toss}, \sim)$

$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 25\%$

---

pretend CH true → see if data under temporary pretend is unlikely; if so, CH looks bad (wrong)

25% not rare | $P(\text{all 9 heart disease deaths smokers} \mid CH)$

$= P(\text{HHHHHHHHHH} \mid \text{IID fair})$

$$= \underbrace{\left(\tfrac{1}{2}\right)\cdots\left(\tfrac{1}{2}\right)}_{9} = \frac{1}{2^9} = \frac{1}{512} \doteq 0.2\% \doteq 0.002$$

(pretend CH true) → (lots extremely unlikely) → CH prob. wrong

---

{T, N}

$\underline{2}\cdot\underline{2}\cdot\underline{2}\cdot\underline{2}\cdot\underline{2} = 2^5 = 32$

↑ 1st 2nd

# of T-s babies
0
1
2
3
4
5

if ELM true

ELM not true

true

$P(1 \text{ or more} \mid \text{ELM}) = \frac{5}{6}$

right:

$1 - \left(1 - \frac{1}{4}\right)^5 \doteq 76\%$

(5)

ok stuff (not dividing $1 > 0$)

okay subsets of S $\mathbb{C}$

dividing $1 > 0$

subsets of 2^S not in $\mathbb{C}$

(mutually exclusive)

A, B <u>disjoint</u> $\leftrightarrow A \cap B = \phi$

$\leftrightarrow$ (T/F prop.) A, B cannot both be true simultaneously

---

A, B <u>independent</u> $\leftrightarrow$ ① Bayesian info. about A doesn't change chances of B & vice versa ② (freq.)

$P(A \overset{\cap}{\underline{\text{and}}} B) = P(A) \cdot P(B)$

ex. of non-independence

pop

$A = \{\text{smoker}\}$

$B = \{\text{risk of heart disease}\}$

$P(B) = \text{small}$

$P(B \mid \underline{\underline{A}}) \uparrow$

A, B dependent

if

$P(A \text{ and } B) = 0$ A, B mut. excl.

and A, B indep $\longrightarrow$

$P(A \text{ and } B) = 0 = P(A) \cdot P(B)$

$\rightarrow$ either $P(A) = 0$ or $P(B) = 0$ or both