this time: foundations ↓

next time: read: Jaynes (2003) (J) ch. 1, 2

AMS 206

9 Jan 18

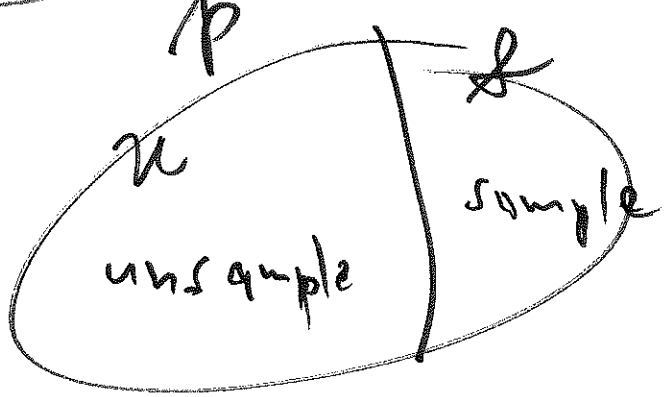
webcast.ucsc.edu | userid ams-206

password: bayesian-data-science

web page
ams206-winter18-01.courses.soe.ucsc.edu

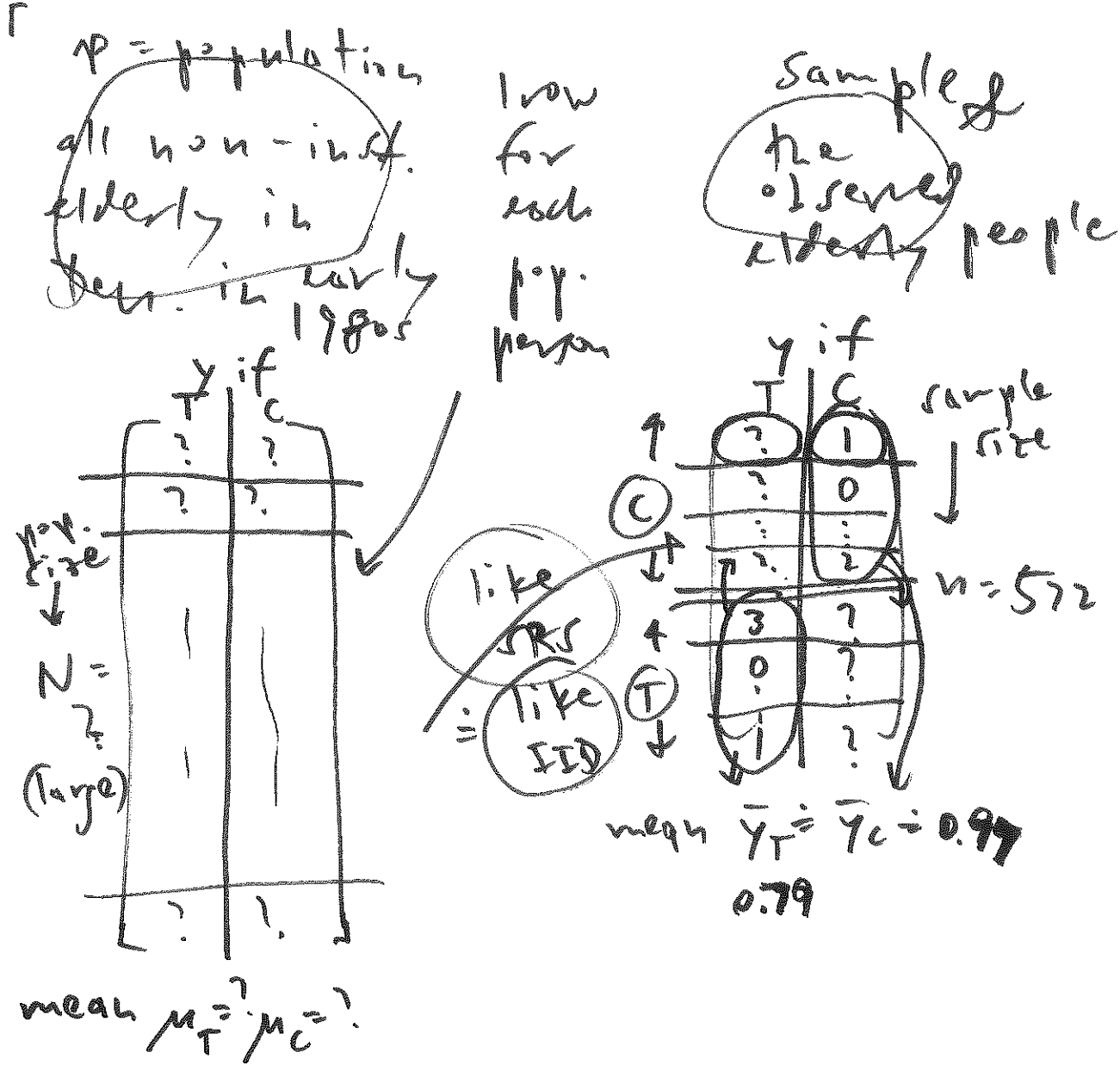
email subject line: ams 206

$\mathcal{P}$ = population



(partition)

a sampling method is representative if it often leads to samples that are similar to $\mathcal{P}$



---

simple method for achieving representativeness:
choose sample at random from pop. (P)

the subsample is all relevant ways

(+)

sampling at random → with replacement (IID) sampling
→ without replacement (SRS) (simple random sampling)

IID = independent identically distributed

SRS is more informative than IID, but math harder with SRS;

if $n \ll N$ ($\ll$ is a lot smaller), SRS ≈ IID

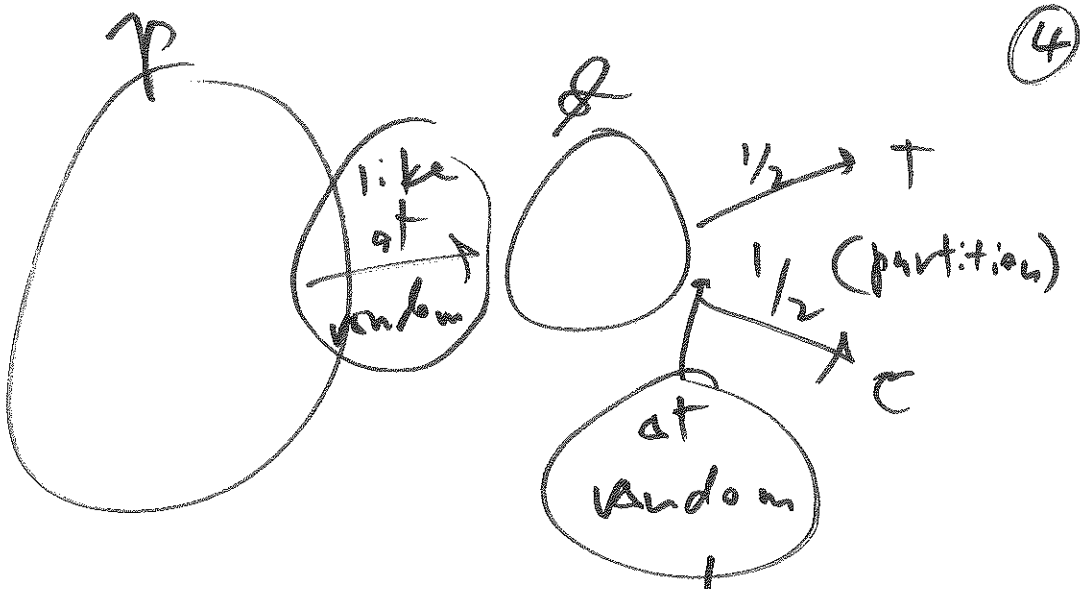
C = control group

not IHGA = control

IHGA = the treatment

T = treatment group

completely randomized design



to promote comparability

(similarity of T, C people in all relevant ways)

outcome:

y = # hospitalizations in 2 yr.

$n = 572$ → $n_T = 285$, $n_C = 287$

why 572?

$n = 572$

sample size determination

if too much uncertainty, get more good data

concentrate replications where your uncertainty is largest

Q1 is $\bar{y}_T = 0.79$ practically significantly different from $\bar{y}_C = 0.97$?

A1 yes; this difference is practically significant

Q2: is $\bar{y}_T = 0.79$ statistically significantly different from $\bar{y}_C = 0.97$?

A2 we don't know yet