

this samples & populations;
 time: histograms

AMS 7
 5 Apr 18

next time: measuring of center & spread

or of ①
 next Mon,

go to discussion section

principle: to decrease your uncertainty (or not completely known) about something unknown to you, gather new good data

population p
 all deer living
 use counts on 31 Mar 18
 column for each variable
 ≈ 800
 row for each element

no
no = 0
yes = 1
no
no

 chronic wasting disease? (CWD)
 binary
 population size
 (1 = yes, 0 = no)
 sum $\sum x_i = \# \text{ deer with CWD}$
 mean parameter $\frac{\sum x_i}{N} = \theta = p = ?$

sample of the observed deer
 CWD?

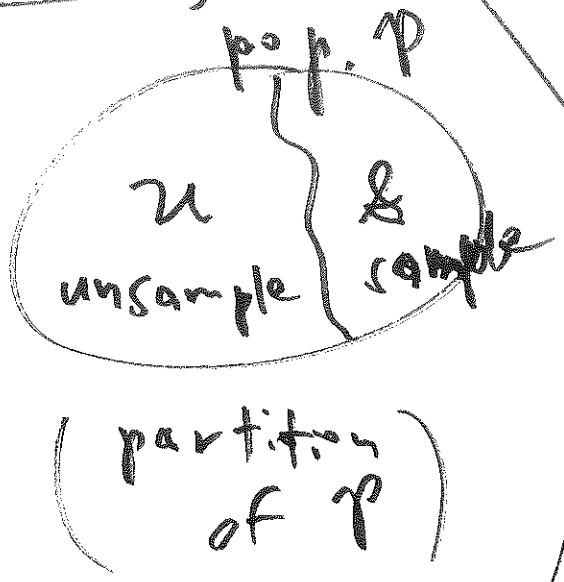
1
5
4
0

 sample size $n = 91$
 our estimate of $\theta = \hat{p}$
 $\sum = 4$
 mean $\frac{\sum}{n} = \frac{4}{91} = 4.4\%$
 square $\frac{\sum x_i^2}{n} = \hat{\theta}^2 = \hat{p}^2$
 by place of Leij
 sampled

at random
 call deer have

graphical & numerical summaries
of data sets: descriptive statistics

goal of sampling → try to make S and \mathcal{U}
as similar possible in



all relevant ways

how to achieve
this goal: choose
 S at random

2 simple
random sampling methods:

- ① at random with replacement
(independent identically distributed (IID))

② at random without replacement ③
(simple random sample (SRS))

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

when $n \ll N$ is a lot smaller than

(sample size) (pop. size)

SRS = IID

perfect similarity & all every time

randomization can't guarantee between S & (in all relevant ways)

but
① the bigger n sets, the more likely that $\hat{\theta}$ & θ are (relevantly similar) ④

② we will learn methods to estimate how often randomization yields

bad samples

unrepresentative $\hat{\theta}$ of θ

since sampling is at random,

$$\hat{\theta} = \hat{p} = \frac{\sum y}{n} = 4.4\%$$

is a good estimate of $\theta = p = \frac{\sum \theta_i}{N}$

how good?

we think θ is
on avd 4.4%, give
or take ?%

variables & the values they take on

variable	possible values
eye color	brown, blue
hair color	brown, black, red, white
success in wage earning	very slow, slow, moderate, fast, very fast

dichotomous ✓

ordered categories
ordinal

qualitative =
(no h-h numeric) =
qualitative

its values do not have

unique places on number line